

ADOC 37BVIIVAV 1S3B

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2004, 09:36:09 ; Search time 47 Seconds
(without alignments)
1644.677 Million cell updates/sec

Title: US-09-825-414-66
Perfect score: 2407
Sequence: 1 MHINSAQQPPGVAMESEFT.....EEGGISAFSRSETPFQLRRL 487

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*

1: /SIDS1/gcgcdata/geneseq/geneseqp-emb1/AA1980.DAT:*

2: /SIDS1/gcgcdata/geneseq/geneseqp-emb1/AA1981.DAT:*

3: /SIDS1/gcgcdata/geneseq/geneseqp-emb1/AA1982.DAT:*

4: /SIDS1/gcgcdata/geneseq/geneseqp-emb1/AA1983.DAT:*

5: /SIDS1/gcgcdata/geneseq/geneseqp-emb1/AA1984.DAT:*

6: /SIDS1/gcgcdata/geneseq/geneseqp-emb1/AA1985.DAT:*

7: /SIDS1/gcgcdata/geneseq/geneseqp-emb1/AA1986.DAT:*

8: /SIDS1/gcgcdata/geneseq/geneseqp-emb1/AA1987.DAT:*

9: /SIDS1/gcgcdata/geneseq/geneseqp-emb1/AA1988.DAT:*

10: /SIDS1/gcgcdata/geneseq/geneseqp-emb1/AA1989.DAT:*

11: /SIDS1/gcgcdata/geneseq/geneseqp-emb1/AA1990.DAT:*

12: /SIDS1/gcgcdata/geneseq/geneseqp-emb1/AA1991.DAT:*

13: /SIDS1/gcgcdata/geneseq/geneseqp-emb1/AA1992.DAT:*

14: /SIDS1/gcgcdata/geneseq/geneseqp-emb1/AA1993.DAT:*

15: /SIDS1/gcgcdata/geneseq/geneseqp-emb1/AA1994.DAT:*

16: /SIDS1/gcgcdata/geneseq/geneseqp-emb1/AA1995.DAT:*

17: /SIDS1/gcgcdata/geneseq/geneseqp-emb1/AA1996.DAT:*

18: /SIDS1/gcgcdata/geneseq/geneseqp-emb1/AA1997.DAT:*

19: /SIDS1/gcgcdata/geneseq/geneseqp-emb1/AA1998.DAT:*

20: /SIDS1/gcgcdata/geneseq/geneseqp-emb1/AA1999.DAT:*

21: /SIDS1/gcgcdata/geneseq/geneseqp-emb1/AA2000.DAT:*

22: /SIDS1/gcgcdata/geneseq/geneseqp-emb1/AA2001.DAT:*

23: /SIDS1/gcgcdata/geneseq/geneseqp-emb1/AA2002.DAT:*

24: /SIDS1/gcgcdata/geneseq/geneseqp-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	2407	100.0	487	22	AAE12603	Pseudomonas syring
2	1869.5	77.7	486	22	AAE12573	Pseudomonas syring
3	1869.5	77.7	486	22	AAE67677	Amino acid sequenc
4	147	6.1	1463	23	AAE20110	Lactobacillus rham
5	146.5	6.1	2122	24	ABU08784	Moraxella catarrha
6	146.5	6.1	2123	22	AAE00701	Moraxella catarrha
7	143	5.9	10431	24	ABU54861	Human CA125 amino
8	140	5.8	2285	20	AAW98149	Bacillus subtilis
9	138	5.7	1229	22	ABB71736	Drosophila melanog
10	138	5.7	1277	22	AAE76536	Corynebacterium gl
11	138	5.7	2012	22	AAE76534	Corynebacterium gl
12	138	5.7	2993	22	AAE92485	C glutamicum prote
13	136	5.7	1026	15	AAE48993	rsaA S-lyase prote
14	132	5.5	1026	18	AAW37490	Caulobacter cresce
15	132	5.5	1026	21	AAE44757	Caulobacter cresce
16	132	5.5	6077	24	ABP95996	White spot syndrom
17	130.5	5.4	2186	24	ABJ19059	Pathogen specific
18	130.5	5.4	2368	22	AAU34139	Staphylococcus aur
19	130.5	5.4	2368	22	AAU36796	Staphylococcus aur
20	130	5.4	1026	17	AAE94014	Caulobacter S-lyase
21	130	5.4	1026	24	AAE34374	Caulobacter cresce
22	129.5	5.4	2076	22	AAU34319	Staphylococcus aur
23	129.5	5.4	2186	22	AAU37320	Staphylococcus aur
24	129	5.4	779	23	ABP65343	Bifidobacterium lo
25	129	5.4	2344	22	AAU37120	Staphylococcus aur
26	129	5.4	6025	22	AAE84939	Shrimp white spot
27	127.5	5.3	2387	22	AAU01183	Rat glutamate tran
28	127.5	5.3	2478	22	AAU34320	Staphylococcus aur
29	127.5	5.3	2478	22	AAU37374	Staphylococcus aur
30	127.5	5.3	2478	24	ABJ19002	Pathogen specific
31	126	5.2	727	22	ABG25530	Novel human diagno
32	125.5	5.2	623	23	AAU12038	Clostridium diffic
33	125.5	5.2	1252	16	AAE80530	B. sphaericus SLP.
34	125	5.2	456	22	ABE64582	Drosophila melanog
35	125	5.2	793	22	AAU27571	Neisseria meningit
36	125	5.2	793	22	AAE10016	N. meningitidis st
37	125	5.2	2086	22	AAU34143	Staphylococcus aur
38	125	5.2	5795	22	AAU37017	Staphylococcus aur
39	124	5.2	628	23	AAE66009	F. necrophorum tru
40	124	5.2	630	24	ABJ26539	Aspergillus fumiga
41	124	5.2	3241	23	AAE66005	F. necrophorum leu
42	124	5.2	3647	11	AAE05041	Filamentous haemag
43	123.5	5.1	2434	22	AAU34339	Staphylococcus aur
44	123.5	5.1	6281	22	AAU37403	Staphylococcus aur
45	123.5	5.1	10498	24	ABJ19119	Pathogen specific

ALIGNMENTS

RESULT 1
AAE12603
ID AAE12603 standard; Protein; 487 AA.
XX
AC AAE12603;
XX
DT 03-JAN-2002 (first entry)
XX
DE Pseudomonas syringae pv. tomato strain DC3000 HopPtoA2 protein.
XX
KW Conserved Effector Loci; CEL; cytosolic; antibacterial; gene therapy;
KW Exchangeable Effector Loci; EEL; disease resistance; transgenic plant;
KW eukaryotic cell death; cancer.
XX
OS Pseudomonas syringae.
XX
PN WO200175066-A2.
XX
PD 11-OCT-2001.
XX
PF 03-APR-2001; 2001WO-US10698.
XX
PR 03-APR-2000; 2000US-194160P.
PR 11-AUG-2000; 2000US-224604P.
PR 17-NOV-2000; 2000US-249548P.
XX
PA (COOR) CORNELL RES FOUND INC.
PA (UYNE-) UNIV NEVADA & COMMUNITY COLLEGE SYSTEM.
PA (UYNE-) UNIV NEBRASKA.
XX
PI Collier A, Alfano JR, Charkowski AO;
XX
DR WPI; 2001-639361/73.
DR N-PSDB; AAD20438.
XX
PT New nucleic acid molecules encoding proteins or polypeptides of
PT Pseudomonas Conserved Effector Loci and Exchangeable Effector Loci
PT genomic sequences, for imparting disease resistance to plants -
XX
PS Claim 8; Page 65-66; 217pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule comprising a
CC nucleotide sequence encoding proteins or polypeptides of Pseudomonas
CC Conserved Effector Loci (CEL) and Exchangeable Effector Loci (EEL)
CC genomic sequences. CEL and EEL DNA are useful for imparting disease
CC resistance to a plant, by transforming a plant cell with the nucleic acid
CC and regenerating a transgenic plant from the transformed plant cell,
CC where the transgenic plant expresses a heterologous DNA molecule under
CC conditions effective to impart disease resistance, or by treating a plant
CC with an isolated protein or polypeptide, by applying the protein or
CC polypeptide in an isolated form or by applying a non-pathogenic bacteria
CC which secretes the protein or polypeptide, under conditions effective to
CC impart disease resistance to the treated plant. CEL and EEL proteins
CC are useful for causing eukaryotic cell death, by introducing a cytotoxic
CC Pseudomonas protein into a eukaryotic cell under conditions effective to
CC cause cell death. CEL and EEL proteins are also useful for treating a
CC cancerous condition, by introducing a cytotoxic Pseudomonas protein into

CC cancer cells of a patient under conditions effective to cause death of
CC cancer cells, and thus treating the cancerous condition. The method
CC further involves administering a targeted DNA delivery system
CC comprising a DNA molecule encoding the cytotoxic Pseudomonas protein,
CC to the patient, where the targeted DNA delivery system delivers the
CC DNA molecule into cancer cells and the cytotoxic Pseudomonas protein
CC is expressed in the cancer cells. The present sequence is
CC Pseudomonas syringae pv. syringae HopPysA homolog protein.

XX SQ Sequence 487 AA;

Query Match 100.0%; Score 2407; DB 22; Length 487;
Best Local Similarity 100.0%; Pred. No. 6.3e-192;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHINSAQQPPGVAMESEFRTASDASIASSSVRSVSTTSCRDLQAITDYLKHHVFAHRS 60
|||||
DB 1 MHINSAQQPPGVAMESEFRTASDASIASSSVRSVSTTSCRDLQAITDYLKHHVFAHRS 60
QY 61 VIGSPDERDALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAF 120
|||||
DB 61 VIGSPDERDALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAF 120
QY 121 ENTFFAASVLYQYQPAINKGDMLATPLKPLPLISGALSGAMDQYGTMMDBARGLHY 180
|||||
DB 121 ENTFFAASVLYQYQPAINKGDMLATPLKPLPLISGALSGAMDQYGTMMDBARGLHY 180
QY 181 LSTSPDKLHDAMAVSVKRHSPALGRQVDMGIAVQTFSAINVVRTVLAALASRPSVGA 240
|||||
DB 181 LSTSPDKLHDAMAVSVKRHSPALGRQVDMGIAVQTFSAINVVRTVLAALASRPSVGA 240
QY 241 VDFGVSTAGLVANAGFGDRMLSVQSRDQLRGAFVLGKDKPKKALSEETDMLDAYKA 300
|||||
DB 241 VDFGVSTAGLVANAGFGDRMLSVQSRDQLRGAFVLGKDKPKKALSEETDMLDAYKA 300
QY 301 IKSASVGAALNAGKRMAGLPIDVATDGLKAVRSLSVATSLTKNGIALAGYAGVSKLQK 360
|||||
DB 301 IKSASVGAALNAGKRMAGLPIDVATDGLKAVRSLSVATSLTKNGIALAGYAGVSKLQK 360
QY 361 MATKNITDSATKAAVSQLSNLVGSVGFAGWTTAGLATDPAVKKAESFIQDKVKSTASST 420
|||||
DB 361 MATKNITDSATKAAVSQLSNLVGSVGFAGWTTAGLATDPAVKKAESFIQDKVKSTASST 420
QY 421 TSYVADQTVKLAKTVKMSGEALISSTGASLRSTVNNLRHSAPEADI EEGGISAFSRSET 480
|||||
DB 421 TSYVADQTVKLAKTVKMSGEALISSTGASLRSTVNNLRHSAPEADI EEGGISAFSRSET 480
QY 481 PFQLRRL 487
|||||
DB 481 PFQLRRL 487

RESULT 2
AAE12573
ID AAE12573 standard; Protein; 486 AA.
XX
AC AAE12573;

Example 2; Page 23; 43pp; English.

SQ Sequence 486 AA;

Matches 381; Conservative 33; Mismatches 71; Indels 1; Gaps 1;

Db 61 PADSADGQAVDVHNAQITALIETRASRLHFEGETPATIADTEAKAEKLDRIATTSGAL 120

Db 480 PFRPMR 485

Db 421 TGYVADQTVKLAKTVKDMGGEATHTGASLRNTVNNLRQRPAREADIEEGTAA-SPSEI 479

Db 480 PFRPMR 485

RESULT 4
AAE20110
ID AAE20110 standard; Protein; 1463 AA

DT 18-JUN-2002 (first entry)

DE Lactobacillus rhamnosus outer membrane protein rompA.

KW Enzyme;flavour; aroma; texture; nutritional; dairy manufacture; therapy;
KW fermentation process; anti-infection; rotavirus infection; heart disease;
KW infantile diarrhoea; lactose digestion; anti-cancer; autoimmune disorder;
KW anti-mutagenesis; immune system modulation; allergy; Helicobacter pylori;;
KW antihypertensive effect; urogenital infection; hepatic encephalopathy;
KW bowel syndrome; endocarditis; transgenic microbe; outer membrane protein;;
KW rompa.

OS	<i>Lactobacillus rhamnosus</i> HN001.
PN	WO200212506-A1.
XX	
XX	14-FEB-2002.
PD	
XX	08-AUG-2001; 2001WO-NZ00160.
PF	
XX	08-AUG-2000; 2000US-0634238.
PR	28-NOV-2000; 2000US-0724623.

PR 28-NOV-2000; 2000US-0724623.

[illegible]

Query Match	6.18;	Score 146.5;	DB 24;	Length 2122;
Best Local Similarity	22.28;	Pred. No. 0.011;		
Matches 111;	Conservative 69;	Mismatches 181;	Indels 139;	Gaps 23

QY	32	RSVSTTSCRDQAITDYLKHHVFAAHRFVSGSPDERDALAHNEQIDALVETR---ANR	88
	::: :	: : : :	
DB	377	KTLITTTGAQTSALTD-----HNIGVONGDGLKVQLAETLITLSLKVVTTENTITANE	4277
QY	89	LYSEGETPATIAETFAKAKEFDRLATTASSAFENTPFAASVLQYMQPAINKGDWILATPL	148
	:	: : : : : :	

```

RESULT 6
AAE00701
ID AAE00701 standard; Protein; 2123 AA.
XX
XX AAE00701;
XX
DT 02-JUL-2001 (first entry)
XX
DE Moraxella catarrhalis outer membrane protein-106 (OMP106).
XX
KW Outer membrane protein-106; OMP106; haemagglutination; vaccine;
KW bacterial infection; immunogen; cytotoxic; antiphotic;
KW passive immunisation.
XX
OS Moraxella catarrhalis.
XX
PN US6214981-B1.
XX
PD 10-APR-2001.
XX
PF 12-NOV-1997; 97US-0968685.
XX
PR 03-MAY-1996; 96US-0642712.
XX
PA (ANTE-) ANTEX BIOLOGICS INC.
XX

```


PI Tucker K, Plosila L, Tillman UF;
XX
DR WPI; 2001-281002/29.
DR N-PSDB; AAD04029.
XX
PT Novel nucleotide sequences encoding Moraxella catarrhalis outer
PT membrane protein-106 polypeptide, useful for diagnosis of bacterial
PT infections and as vaccine against Moraxella catarrhalis infection of
PT mammals -
XX
PS Claim 7; Column 53-64; 49pp; English.
XX
CC The present sequence is haemagglutinating Moraxella catarrhalis outer
CC membrane protein-106 (OMP106). The OMP106 is used as a therapeutic
CC and prophylactic vaccine against M. catarrhalis infections of mammals.
CC It is used for diagnosis of bacterial infections and as reagents for
CC clinical or medical diagnosis of M. catarrhalis infections and for
CC scientific research on the properties of pathogenicity, virulence and
CC infectivity of M. catarrhalis. It is also used as a probe to identify
CC the presence of M. catarrhalis in biological specimens and to identify
CC other bacteria that encode a polypeptide related to M. catarrhalis
CC OMP106. OMP106-derived polypeptides are used as ligands to detect
CC antibodies elicited in response to M. catarrhalis infections and also
CC as immunogens for inducing M. catarrhalis-specific antibodies which are
CC useful in immunoassays to detect M. catarrhalis in biological specimens.
CC Cytotoxic antibodies are useful in passive immunisations against
CC M. catarrhalis.
XX
SQ Sequence 2123 AA;

Query Match 6.1%; Score 146.5; DB 22; Length 2123;
Best Local Similarity 22.2%; Pred. No. 0.011;
Matches 111; Conservative 69; Mismatches 181; Indels 139; Gaps 23;

QY 32 RSVTTSRCRDQAITDYLKHNFAAHRFSVIGSPDERDALAHNEQIDALVER---ANR 88
Db 377 KTLITTTGAQTSALT-----HNIGVVQNGDGLKQQLAETLTSLKMTTENLTANE 427
QY 89 LYSEGETPATIAETFAKAEKFDRLATTASSAFENTPFAASVLQYKQPAINKGDWLATPL 148
Db 428 KVTVGKT-----RL-TTDKIGFTND---MNGIDESKPYLDKDTGIHAGG 467
QY 149 KPLTPPLISGAL-----SGAMDYVGTQMDRARGLHYLSTSPDKLHDA 191
Db 468 QKITTKLTAGVDDDAATYQGLKRVNQTAESALQFTVKVDKNGND-----ANDS 517
QY 192 MAVSVKRHSPALGRQVVDM-----GIAVQTFSAINVRVTLAPALASRPVGVGAVDFGV 245
Db 518 KIITVGKNNKPDGTQVNTLKLKGENGVDTT-----ETNGTVTFGL 558
QY 246 STAGGL-VANAGFGDRMLSVQ-----SRDQLRGAFVLGMD---KEPKALISEET----DW 294
Db 559 NQNNGLTVGNSTLNDGSLVKNINSNKQIQVGADGITFTDISNSKPGAGIENTTRITRDG 618
QY 295 LDAYKAIKSASYSGAALNAGKRNAGLPDVATD---GLKAVRSLSVSAATSLTKNGLALAG 351
Db 619 I-----GFAANNTGSLDANKPRLPTGINAGGKELTNVQSAINPATNG----- 660

QY 352 YAGVSKIQKMATKNITDSATKAAVSQSLNVGSVGFAGWTTAGLATDPAV-KKAESFIQ 410
Db 661 -GQLDFWNRLSTANTEKSSGSAATIKDLXNLSQVPLTFAG-----DTGPNVTKKLGEILK 713
QY 411 DK-VKSTASST-----SYVADQ-----TVKLAKTVKOMSGEALISSTGASLRSTVNLRHR 460
Db 714 VKGKTTADDLTKNNIGVADSTDNSLTVKLAKTLSLDDAVNTKTLTASDKVTYDSGAN- 772
QY 461 SAPEADIEEGGISAFSRSET 480
Db 773 ---TAKIQNGDLT-FSKQNT 788

RESULT 7
ABU54861
ID ABU54861 standard; Protein; 10431 AA.
XX
AC ABU54861;
DT 12-MAR-2003 (first entry)
XX
DE Human CA125 amino terminal extension.
KW Human; CA125; protein repeat; chromosome 19q 13.2; amino terminal domain;
KW amino terminal extension; carboxy terminal domain; vaccine; cancer;
KW ovarian cancer; carcinoma.
XX
OS Homo sapiens.
XX
PN WO200283866-A2.
XX
PD 24-OCT-2002.
XX
PF 12-APR-2002; 2002WO-US11734.
XX
PR 17-APR-2001; 2001US-284175P.
PR 19-JUN-2001; 2001US-299380P.
PR 27-SEP-2001; 2001US-0965738.
PR 21-DEC-2001; 2001US-345180P.
XX
PA (UYAR-) UNIV ARKANSAS.
XX
PI O'Brien T, Beard J, Underwood L;
XX
DR WPI; 2003-093013/08.
DR N-PSDB; ABX72626.
XX
PT New CA125 molecules, useful as a gold standard for detecting and
PT monitoring the presence of CA125 antigen which can be used for
PT diagnosing, monitoring or treating patients with cancer or for
PT developing vaccine against cancer -
XX
PS Claim 1; Page 158-169; 694pp; English.
XX
CC The invention relates to a CA125 protein comprising: (a) an extracellular
CC amino terminal domain; (b) an amino terminal extension; (c) a multiple

Sequence 10431 AA;

Matches 113; Conservative 87; Mismatches

Dib

9b

ID AAW

DE

KM

50

PN

PD

PF

PR

PA

Id

DR

14

PS

8

CC agricultural applications, and advantageous strain improvements
CC based on genetically engineering a Gram-positive microorganism to
CC delete, underexpress or overexpress the enzyme. Due to overall
CC relatedness of the enzyme with *Pseudomonas* lasA protein, YOMI
CC appears to be a member of the M23 metalloprotease family. The
CC metalloprotease can be used in claimed cleaning compositions,
CC animal feed and compositions for the treatment of textiles. It may
CC also be used for peptidic hydrolysis, waste treatment and for
CC cleaving recombinant fusion proteins. Expression vectors
CC comprising a nucleic acid (see AAX24980) encoding the metalloprotease
CC and host cells are claimed. Gram-positive cells in which YOMI is
CC inactivated (by gene mutation or deletion) are used for production
CC of heterologous proteins, especially enzymes, hormones, growth
CC factors and cytokines.

SQ Sequence 2285 AA;

Query Match	5.88;	Score 140;	DB 20;	Length 2285;
Best Local Similarity	20.58;	Pred. No. 0.041;		
Matches 128;	Conservative 80;	Mismatches 224;	Indels 192;	Gaps 26;

QY 15 MESERTASDASSVRSVSTTSCRDQAITDYLKHVEAAHRSV-IGSPDERDALA 73

Db 251 I ELYQRQAQVNVQN L N T R Y G S S M G S S N R Q A V Q D Y L N --- A V N S L N V S T G S N N I R S Q I Q S

74 HNEQIDAL---VETRANRLYSEGETPATIAETF-----AKA 106

Db 307 LNMQFRELASNAQTANQASSFG--AELTQTFKSMSTYLLISGSLFYGAISGLKEMV\$QA 363

107 EKEDRLATTASSAFENTPFAASVLQYMQPAINKGDWLATPLKPLTPLISGALSAMDQV 166

Db 364 I E I D T L M T N I R R V M N E P D Y K Y N E L L -- Q E S I D I G D T L S N K I T D I L Q M T G D F G R M G F D E S 420

167 GTKMDRARGDLHYLS-TSPDKLHDAMAVSVKRRHSPALGRQV-----VDMGIADVQTF 217

Db 421 ELSTLTQAQVLQNVSDLTPDDTVNTLTAAMLNENIAANDSISIADKLINEVDNNYAVTTL 480

218 SALNVVRTVLAPALASRPSVQGAVDFGVSTA-----GGLVAN-----AGFGDRMLLS 263

Db 481 DIANSIRK--AGSTASTFGVELNDLIGYTTAIASTTRESGNIVGNSLKTIFARIENNOSS 538

264 VQSRDQLRGAFVLGKKOKEPKALSE---ETDWL-DAYKAIKSASYSG----- 308

Db 539 IKALEQIGISVKTAGGEAKSASDLISEVAGKWDLTSDAQKQNTSIGVAGIYQLSRFNAMM 598

QY 309 ---BALNAGK--- 315

Db 599 NNFSLAQNAKTAANSTGSAMSEQQKYADSLQARVNNKLNNFTFAIAASDAFISDGLIE 658

316 --RMAGLPDVATDGLKAV---RSLVSATS-----LTKNGLALAGG-YAGVSKL--QKMA 362

Db 659 FTQAAGSLNASTGVIKSVGFPLPLAAVSTATLLSKNTRTLASSLLIGTRAMQETLA 718

QY 363 THNITDSATKAVSQ-----LSNLVGSVGEAGWTTAGLATDPA-VKKA----- 405

Db 719 TAGLEAGMTTAAVASRVLKTLRGLVSTLVGGAFAALGNALSLISSFAEAKKAKADDEE 7/8

```

QY 406 -----ESFIQ-----DKKSTASSTTSVADQ-----TVKAKT-----VK 436
Db 779 QSQOQTNVEAITTNKSDTDKLIQYKELQVKESRSLTSDBOEYIADVTQQLAQTFPALVK 838
QY 437 --DMSGEAIISSTGASLSTSTVNNLR 458
Db 839 GYDSQGNAILKTNKELEKAIENTK 862

```

RESULT 9

ID ABB71736 standard; Protein; 1229 AA.

AC ABB71736;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 42000.

KW *Drosophila*; developmental biology; cell signalling; insecticide;

XX

XX

✕✕

PR 11-JUL-2000; 2000US-0614150-

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW,

DR WPI; 2001-656860/75.

1. ☒ 2. ☒ 3. ☒ 4. ☒ 5. ☒ 6. ☒ 7. ☒ 8. ☒ 9. ☒ 10. ☒ 11. ☒ 12. ☒ 13. ☒ 14. ☒ 15. ☒ 16. ☒ 17. ☒ 18. ☒ 19. ☒ 20. ☒ 21. ☒ 22. ☒ 23. ☒ 24. ☒ 25. ☒ 26. ☒ 27. ☒ 28. ☒ 29. ☒ 30. ☒ 31. ☒ 32. ☒ 33. ☒ 34. ☒ 35. ☒ 36. ☒ 37. ☒ 38. ☒ 39. ☒ 40. ☒ 41. ☒ 42. ☒ 43. ☒ 44. ☒ 45. ☒ 46. ☒ 47. ☒ 48. ☒ 49. ☒ 50. ☒ 51. ☒ 52. ☒ 53. ☒ 54. ☒ 55. ☒ 56. ☒ 57. ☒ 58. ☒ 59. ☒ 60. ☒ 61. ☒ 62. ☒ 63. ☒ 64. ☒ 65. ☒ 66. ☒ 67. ☒ 68. ☒ 69. ☒ 70. ☒ 71. ☒ 72. ☒ 73. ☒ 74. ☒ 75. ☒ 76. ☒ 77. ☒ 78. ☒ 79. ☒ 80. ☒ 81. ☒ 82. ☒ 83. ☒ 84. ☒ 85. ☒ 86. ☒ 87. ☒ 88. ☒ 89. ☒ 90. ☒ 91. ☒ 92. ☒ 93. ☒ 94. ☒ 95. ☒ 96. ☒ 97. ☒ 98. ☒ 99. ☒ 100. ☒

PT genes from *Drosophila* and for elucidating cell signalling and cell-cell

✕✕

✕✕✕

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

Pt transformation -
XX
PS Claim 20; Page 242-246; 1119pp; English.
XX
CC AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane
CC construction and membrane transport (MCT) proteins given in AAB76510 to
CC AAB76847. The MCT nucleic acids and proteins are useful in the
CC identification of microorganisms which can be used to produce fine
CC chemicals, for modulating fine chemical production in C. glutamicum or
CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or
CC identification of C. glutamicum or related bacteria, as reference points
CC for mapping C. glutamicum genome, and as markers for transformation.
CC AAF68082 and AAF68082 represent sequencing primers which are used in an
CC example from the present invention.

XX
SQ Sequence 1277 AA;
Query Match 5.7%; Score 138; DB 22; Length 1277;
Best Local Similarity 21.7%; Pred. No. 0.027;
Matches 119; Conservative 85; Mismatches 219; Indels 126; Gaps 23;

QY 7 AQQPPGVAMSESFRTASDASLASSSVRSVSTSCRDLQAITDYLKHVFAAHFVSIGSPD 66
Db 313 AHTAPDVLVGRAMPVFAAVKSAVI PGTFDASVVEGMLSLVLEHHI-----VLKSDV 365
QY 67 ERDMLAHNEQIDALVETRANRLY-----SEGETPATIAETFA-KAEKFDRLATAS 117
Db 366 PTGALKVSATADEVVDITDLGRLLVIRAEIADAEGLIATLAERFAIRGRKGNVAARTNT 425
QY 118 SAFE---NTPFAASVLQYMQP-----AINKGWLATPLKPLTPL---ISGALSGAMD 164
Db 426 SALPTTVDTPRSARAVATVVAPESMRPFAVISGD-----RNPIHVS DVASLAGJPG 477
QY 165 QVGTQKMDRAGDLHYLSTSPDKLHDAMAVSVKRH-----SPALGRQVWDGI---AVQT 216
Db 478 VIVHGMMTSALGELIAGAFNDEQIQTPAKKVEYTATMLAPVLPGEIEIEFVSERSAVDN 537
QY 217 FSALNVVRTVLA PALASRPVQGA VDFGVSTAGLVA--NAGFGDRMLSVQSRDQLRGGA 274
Db 538 RPKMGEVRTVTA-----TVNGNL---VLTATAVVAAPSTFYAFPGQGIQSG----- 581
QY 275 FVLGKKDKPKKALISEETDMLDAYKAIKSASYSGAALNAGKRMAGLP L DVAITDGLKAVR- 333
Db 582 --MGHEARRNSQAARAIVDRADAHTRNK-----LGFSTIVEI VENNPREVTVA GEKEFHP 633
QY 334 -SLVSATSLTNGGLAAGYAGYSKLQKVA TKNITDSATKA VSQLSNLVGSVGVFAGMT 392
Db 634 DGVLYLTQFTQVGMATL-GVAQIAEMREAHALNQRAY FAGHSVGEYNALAAVAGVLSLES 692
QY 393 T-----AGLATDPAVKKAESFIQ-----DKVKSTASSTTSYVADQTVKLAKTVKD 437
Db 693 VLEIIVYRGLTMRHLVDRDENGLSNYALALRPNKMGGLTADNVF DYVA-----SVSE 744
QY 438 MSGE-----AISTGASLRSTVNNLHRSAPEADIEE--GGISAF---SRS 478
Db 745 ASGEFLEIVNNVNLAGLOYAVAGTQAGL-----AALRADVENRAPGQRAFFILIPGI 794

QY 479 ETPFQLRRL 487
Db 795 DVPFHSSKL 803

RESULT 11
AAB76534
ID AAB76534 standard; Protein; 2012 AA.
XX
AC AAB76534;
XX
DT 11-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum MCT protein SEQ ID NO:50.
XX
KW Corynebacterium glutamicum; brevibacterium lactofermentum; MCT;
KW membrane construction and membrane transport protein; petroleum spill;
KW hydrocarbon degradation; gram positive aerobic bacterium; marker;
KW identification; microorganism; fine chemical production; transformation;
KW genome mapping; genetic engineering.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100805-A2.
XX
PD 04-JAN-2001.
XX
PE 23-JUN-2000; 2000WO-IB00926.
XX
PR 25-JUN-1999; 99US-0141031.
PR 08-JUL-1999; 99DE-1031454.
PR 08-JUL-1999; 99DE-1031478.
PR 08-JUL-1999; 99DE-1031563.
PR 09-JUL-1999; 99DE-1032122.
PR 09-JUL-1999; 99DE-1032124.
PR 09-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032128.
PR 09-JUL-1999; 99DE-1032180.
PR 09-JUL-1999; 99DE-1032182.
PR 09-JUL-1999; 99DE-1032190.
PR 09-JUL-1999; 99DE-1032191.
PR 09-JUL-1999; 99DE-1032209.
PR 09-JUL-1999; 99DE-1032212.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032229.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032927.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 14-JUL-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 27-AUG-1999; 99DE-1040830.
PR 27-AUG-1999; 99DE-1040831.
PR 27-AUG-1999; 99DE-1040832.
PR 27-AUG-1999; 99DE-1040833.

PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041395.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042078.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042088.
XX
PA (BADI) BASF AG.
XX
PI Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;
XX
DR WPI; 2001-071486/08.
DR N-PSDB; AAF67767.
XX
XX
PT Corynebacterium glutamicum nucleic acids encoding membrane construction
PT and membrane transport proteins or their portions, useful for typing or
PT identifying C. glutamicum or related bacteria, and as markers for
PT transformation -
XX
PS Claim 20; Page 224-231; 1119pp; English.
XX
CC AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane
CC construction and membrane transport (MCT) proteins given in AAB76510 to
CC AAB76847. The MCT nucleic acids and proteins are useful in the
CC identification of microorganisms which can be used to produce fine
CC chemicals, for modulating fine chemical production in C. glutamicum or
CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or
CC identification of C. glutamicum or related bacteria, as reference points
CC for mapping C. glutamicum genome, and as markers for transformation.
CC AAF68082 and AAF68082 represent sequencing primers which are used in an
CC example from the present invention.
XX
SQ Sequence 2012 AA;

Query Match 5.7%; Score 138; DB 22; Length 2012;
Best Local Similarity 21.7%; Pred. No. 0.051;
Matches 119; Conservative 85; Mismatches 219; Indels 126; Gaps 23;
QY 7 AQQPPGVAMESFRTASDASLASSSVRSVSTSCRDLOAITDYLKHHVEAAHRSVIGSPD 66
Db 995 AHTAPDVLVGRAMPAVFAVKSAVIPGTDASVVEGMLSLVHLEHHT-----VLKSDV 1047
QY 67 ERDALAHNEQIDALVERANRLY-----SEGETPATIAETFA-KAEKEDRLATTS 117
Db 1048 PTDGALKVSATADEVDTDLGRLYIVRAEIDAEENLIATLAERFAIRGRKGNAAVARTNT 1107
QY 118 SAFE--NTPFAASVLYQMOP-----AINKGWLATPLKPLTPL---ISGALSGAMD 164
Db 1108 SALPTTVDTPRSARAVATVVAPESMRPFAVISGD-----RNPITHVS DVAASLAGLPG 1159
QY 165 QVGTKNMDRARGDLHYLSTSPDKLHDAMAVSVKRH-----SPALGRQVVDWGI---AVQT 216
Db 1160 VIVHGMTSAIGELIAGAFNDEQIQTPAAKVVEYTATMLAPVLPGEIEI EFSVERSAVDN 1219
QY 217 FSAINVVRTVLAPALASRPSVQGAVDGVS TAGGIVA--NAGFGDRMLSVQSRDQLRGA 274
 : ||||| | : | | : | | : || : : || :

Db 1220 RRGMEVARTVTA-----TVNGNL---VLTATAVVAAPSTFYAFPGGCIQSQC----- 1263
QY 275 FVLGMRKKEPKALSEETBMLDAYKAIKSASYSGAALNAGKRNAGLPDLVATDGLKAVR- 333
 :||: : | | | | : : : | : | |
Db 1264 --MGMEARRNSQARA IWRADAHTRNK-----LGE SIVEIENNPREVTVAGEKEFHP 1315
QY 334 -SLVSATSLTNGIALAGGYAGVSKIQRMATKNITDSATKAAVSOLSNIWGVGFAGWT 392
 : : | | : | | | : : : : | : | : :
Db 1316 DGVLYLTQFTQVGNATL-GVAQI AEMREAHALNQRAYFAGHSVGEYNALAAVAGVLSLES 1374
QY 393 T-----AGLATDPAVKKAESFIQ-----DKVNSTASSYTSYVADQTVKLAKTVKD 437
 || | : | : : : | : | : | | : | :
Db 1375 VLEIVYRRGLTMRHLVD RDENGLSNVALAALRPNKMGLTADNVFDYVA-----SVSE 1426
QY 438 MSGE-----AISTGASLRSTVNNLRHRSAP EADIEE--GGISAF---SRS 478
 ||| | : | | | : | | | : | | |
Db 1427 ASGEFLEIVNYNLAGLQYAVAGTQAGL-----AALRAVDENRAPGQRAFILIRGI 1476
QY 479 ETPFQLRRL 487
 : || : |
Db 1477 DVPFHSSKL 1485

RESULT 12
AAG92485
ID AAG92485 standard; Protein; 2993 AA.
XX
AC AAG92485;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 6239.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PE 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochial K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR N-PSDB; AAH67704.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying

PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT _
XX
PS Claim 17; SEQ ID NO: 6239; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.

XX
SQ Sequence 2993 AA;

Query Match 5.7%; Score 138; DB 22; Length 2993;
Best Local Similarity 21.7%; Pred. No. 0.089;
Matches 119; Conservative 85; Mismatches 219; Indels 126; Gaps 23;

QY 7 AQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVEAAHRSVIGSPD 66
Db 1025 AHTAPDVLVGRAMPVAFVAKSAVPGTDSASVVEGMLSLVHLEHH-----VLKSDV 1077
QY 67 ERDAAAHNEQIDALVERANRLY-----SEGETPATIAETFA-KAEKEDRLATTAS 117
Db 1078 PTDGALKVSATADEVDTDLGRLIVRAEIDAEGNLIAITLAERFALRGRKGAVARNT 1137
QY 118 SAFE---NTPFAASVLOYMOP-----AINKGMLATPLKPLTPI---ISGALSAMD 164
Db 1138 SALPTTVDTPRSARAVALVAPESMRPFAVISGD-----RNPITHSVDAASLAGLPG 1189
QY 165 QVGTMMMDRARGDLHYLSTSPDKLHDAMAVSVKRH-----SPALGRQVVDMGI---AVQT 216
Db 1190 VIVHGMTSAIGELIAGAFNDEQIQTPAAKVVEYTAATMLAPVLPGEIEIEFSVERSAYDN 1249
QY 217 FSAINVVRTVLAAPALASRPSVQGVDFGVSTAGLVA--NAGFGDRMLSVQSRDQLRGGA 274
Db 1250 RPKMGEVRTVTA-----TVNGNL---VLTATAVVAAPSTFYAFPQGQIQSQG----- 1293
QY 275 FVLGMDKEPKKALSEETDMLDAYKAIKSASYSGAALNAGKRMAGLPDVATDGLKAVR- 333
Db 1294 --MGHEARRNQAARATWDRADAHTRNK-----LGEFIVEIVENNPREVTVAGEKFFHP 1345
QY 334 -SLVSATSLTKNGLALAGYAGVSKLQKATKNITDSATKAASQSLNLVGSVGVFAGMT 392
Db 1346 DGVLYLTQFTQVGMATL-GVAQIAEMREAHALNQRAVFAGHSVGEVNALAAVAGVLSLES 1404
QY 393 T-----AGLATDPAVKKAESFIQ-----DKVKSTASSTTSYVADQTVLAKTYKD 437
Db 1405 VLEIYYRRGLTMHRLVDRDENGSLNYALALRPNKMGLTADNVFDYVA-----SVSE 1456

QY 438 MSGE-----AISSTGASLRSTVNNLRHRSAP EADIEE--CGISAF---SRS 478
Db 1457 ASGEFLEIWNYNLAGLQYANVAGTQAGL-----AALRADEVENRAPGQRAFILIPGI 1506
QY 479 ETPEQLRRL 487
Db 1507 DVPEFHSSKL 1515

RESULT 13
AAR48993
ID AAR48993 standard; Protein; 1026 AA.

AC AAR48993;
DT 14-SEP-1994 (first entry)
XX
DE rsaA S-lyaeer protein.

KW C. crescentus; rsaA; paracrystalline; S-layer; protein; heterologous;
KW cellulase; xylase; metallothionein; restriction site;
KW reading frame; fusion protein; bioreactor; toxic metal; sewage;
KW waste water; wood pulp suspension; cell surface; vaccine; fish.
XX
OS Caulobacter crescentus.

XX
PN CA2090549-A.
XX
PD 10-DEC-1993.

XX
PF 26-FEB-1993; 93CA-2090549.
XX
PR 09-JUN-1992; 92US-0895367.

XX
PA (UYBR-) UNIV BRITISH COLUMBIA.

XX
PI Bingle WH, Smit J;

XX
DR WPI; 1994-066249/09.
DR N-PSDB; AA057972.

XX
PT Prodn. of heterologous polypeptides in bacteria, partic.
PT Caulobacter - by expression of a fusion prod. of the polypeptide
PT sequence and a bacterial S-layer protein gene

XX
PS Claim 17; Fig 6; 27pp; English.

XX
CC This sequence is encoded by the C. crescentus rsaA gene and represents
CC the paracrystalline S-layer protein. The rsaA gene was used in the
CC production of the heterologous protein of the invention. The
CC heterologous protein is produced by cloning a polypeptide coding
CC sequence, eg. cellulase, xylase or a metallothionein, into a
CC restriction site within the rsaA gene which preserves the rsaA reading
CC frame and expressing the fusion sequence in Caulobacter. This S-
CC layer protein bacterial system can be used in bioreactors, eg. to
CC bind toxic metals in sewage waste water etc. or for the treatment of

SQ Sequence 1026 AA;

Matches 102; Conservative 75; Mismatches 165; Indels 140; Gaps 23;

Db 42 SDAALTNLTKLNSTTAVAIQT-----YQFTGVAPSAAG-----LDFL 81

Db 82 V D S T N T N D I N D A Y Y S K - - - - - F A Q E N R F I N F S I N L A T G A G A G - - A T A F A A A Y T G 1 2 2 9

Db 130 VSYAQT VATATYDKI GNAVAT-----AAGVDVAAVAFLSRQANIDYLT--- 173

Db 174 -----AFVRANTPTAAADIDLAVKALIGTILNATVSGIGGYATATAAMINDLSD 225

Db 226 GALSTDNAAGVNLFTAYPPSSGVSGSTLSLTRDTLTGTANDTFVAGEVAGAATLTVGD 285

Db 286 TISGGAGTDLINWVOA-----AAVTA-----LPTGVTISGIETM-NVTSGAAL 327

Db 328 TTTSSGVLTGTAINTNTSGAOTVTAGAGONTATTAAOAAANNVAVDGANVTVAS 384

385 T E V T S G T T V G ----- A N S A A S G T V S V A N S T T T T G A T A - V T T G G T A V T T A O T A G N A V N 438

439 FT 440

DT 20-APR-1998 (first entry)

05 *Caulobacter crescentus*.

PD 18-SEP-1997.

PR 12-MAR-1996; 96US-0614377.

PI Bingle WH, Nomellini JF, Smit J;

XX

XX

CC example of the present invention. A new DNA construct has been developed

CC an ideal system for presentation of antigens, at high level.

50 Sequence 1026 AA;

Matches 101; Conservative 75; Mismatches 166; Indels 140; Gaps 23;

Db 42 SDAALTNLKLNVNSTAVAIQT-----YQFTGVAPSAAG-----LDLFL 81

02 82 VETLAN-----RLYSEGETPATIAETFAKAEKF-----DRLATTASSAFENTPFAA--- 128

Db 82 VDSSTNTNDLNDAYYSK-----FAQENRFINFSINLATGAGG--ATAFAAAAYTG 129
QY 129 -SVLQYMQPAINK--GDMLATPLKPLTPLISGALSGAMDQVGTQMMDRARGDLHYLSTSP 185
Db 130 VSYAQTVATAYDKIIGNAVAT-----AAGVDVAAAVALFSRQANIDYLT--- 173
QY 186 DKLHDAMAVSVKRRHSPALGRQVVDMGIAVQTFSAINVRIVL---APALASRPSVQGAVD 242
Db 174 -----AFVRANTPFTAAADIDLAVKALIGTILNAATVSGIGGYATATAAMINDLSD 225
QY 243 FGVST--AGGI-----VANAGGDRMLSVOQS-RDQLRG-----GAFVLGKMD 281
Db 226 GALSTDNMAAGVNLFTAYPSSGVSSTLSLTGTDTLTGTANNDFVAVAGEVAGAAATLTIVGD 285
QY 282 KEPKALISEETDMLDAYVKAIKSASYSGALNAGKRWAGLPDVAITDGLKAVRSLSVATSL 341
Db 286 TISGGAGTDVLNMQA-----AAVTA-----LPTGVTISGIETM-NVTSGAAL 327
QY 342 TRNGIALAGGYAGVSKL-----QKMATKNITDSATKAAVSQLSNL-VGS 384
Db 328 TLN---TSSGVTGLTALNTNTSGAAQTVTAGAGQNLTAATAQAANNVAVDGRANVTVAS 384
QY 385 VGVFAGWTTAGLATDPAVKKAESFIQDKVKSTASSTSYADQTVKLAKTVKDMSGEALS 444
Db 385 TGVTSGETTVG-----ANSASGTVSVANSSTTTTGALA-VTGGTAVTVAQTAGNAVN 438
QY 445 ST 446
Db 439 TT 440

RESULT 15
AAV44757
ID AAV44757 standard; Protein; 1026 AA.
AC AAV44757;
XX
DT 04-MAY-2000 (first entry)
XX
DE Caulobacter crescentus surface layer protein.
XX
KW Surface layer protein; S-layer secretion signal; antibiotic; vaccine; recombinant fusion protein cleavage; enzyme; protein polymer; antibacterial enzyme; foodstuff.
XX
OS Caulobacter crescentus.
XX
FH Key Location/Qualifiers
FT Cleavage-site 692..693
FT /note= "Aap-Pro dipeptide present in S-layer secretion signal sequence. It is a site where a fusion protein comprising a target protein and the secretion signal is cleaved"
XX
PN WO200004170-A1.
XX

PD 27-JAN-2000.
XX
PF 14-JUL-1999; 99WO-CA00637.
XX
PR 14-JUL-1998; 98CA-2237704.
XX
PA (UYBR-) UNIV BRITISH COLUMBIA.
XX
PI Smit J;
XX
DR WPI; 2000-182434/16.
XX
DR N-PSDB; AAZ50079.
XX
PT Cleavage of Caulobacter produced recombinant fusion proteins useful for producing vaccine peptides -
XX
PS Example 1; Pages 21-23; 33pp; English.
XX
CC The patent discloses a method for cleaving a recombinant fusion protein which is produced by Caulobacter and consists of Caulobacter surface layer (S-layer) protein (containing the C-terminal secretion signal) and a target protein heterologous to Caulobacter. The cleavage of target protein from the S-layer protein is carried out under mild acid conditions so that cleavage occurs at aspartate-proline dipeptide site without solubilising the protein.
CC The cleavage is accomplished while the fusion protein is in an insoluble aggregate form which facilitates purification of the protein. The method is useful for producing pure proteins including recombinant human and animal therapeutic antibiotic and vaccine peptides, enzymes, protein polymers, and antibacterial enzymes for foodstuffs.
CC The present sequence is a S-layer protein from C. crescentus.
CC The S-layer secretion signal, corresponding to the C-terminal portion of the protein from amino acid 690 onwards, is fused with a target sequence for construction of a recombinant fusion construct which is expressed in CC Caulobacter.
XX
SQ Sequence 1026 AA;

Query Match 5.5%; Score 132; DB 21; Length 1026;
Best Local Similarity 21.0%; Pred. No. 0.063;
Matches 101; Conservative 75; Mismatches 166; Indels 140; Gaps 23;
QY 22 SDASLASSSVRSVSTSCRDLQAITDYLKHHVFAHRRFSVIGSPDERDALAHNEQIDAL 81
Db 42 SDAALTLTKLVNSTTAVAIQT-----YQFTGVAPSAAG-----LDFL 81
QY 82 VETRAN-----RUYSEGETPATIAETFAKAEF----DRLATYASSAENTPEAAA--- 128
Db 82 VDSSTNTNDLNDAYYSK-----FAQENRFINFSINLATGAGG--ATAFAAAAYTG 129
QY 129 -SVLQYMQPAINK--GDMLATPLKPLTPLISGALSGAMDQVGTQMMDRARGDLHYLSTSP 185
Db 130 VSYAQTVATAYDKIIGNAVAT-----AAGVDVAAAVALFSRQANIDYLT--- 173
QY 186 DKLHDAMAVSVKRRHSPALGRQVVDMGIAVQTFSAINVRIVL---APALASRPSVQGAVD 242
Db 174 -----AFVRANTPFTAAADIDLAVKALIGTILNAATVSGIGGYATATAAMINDLSD 225

```

QY      243 FGVST--AGGL-----VANA GFDRMLSVQS -RDQLRG-----GAFVLGAKD 281
      :| | | : :| | : | | |
Db      226 GALTSDNAAGVNFETAYPSSGVSSTLSLTGTDTLTGRANNDFVAGEVAGAALTITVGD 285
      :| | | : :| | : | | |
QY      282 KEPKALSEETDWILDAYKAIKSASYSGAALNAGKRMAGLPIDVATDGLKAVRSLSVATSL 341
      | : : | : | | | | | : :
Db      286 TISGAGCTDVLNNVQA-----AAVTA-----LPTGVTTISGIEFM-NVTSGAAT 327
      :| | | : :| | : | | |
QY      342 TRKGLAAGGYAGVSKL-----QKMATKNITDSATKAAVSQLSNI-VGS 384
      | | : : | : : | : :| | : | |
Db      328 TLN---TSSGVTGLTALNTNTSGAAQVTYAGCQNLTAATTAAQANNVAVDGRANVTVAS 384
      :| | : :| | : | : :| | : | |
QY      385 VGVFAGWTTAGLATDPAVKRAESFIQDKVKSTASTTSVADQTVKLAKTVKDMSEGAIS 444
      | | : | | | | | : | : : :| | : | | | : | | : | | :
Db      385 TGVTSQTTTVG-----ANSAAAGTVSVSVANSSTTTGALV-YTGGTAVTVAAQTAGAVNN 438
      :| | | : :| | : | | |
QY      445 ST 446
      :|
Db      439 TT 440

```

Search completed: January 6, 2004, 09:44:21
Job time : 49 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2004, 09:38:14 ; Search time 20 Seconds
(without alignments)
2341.707 Million cell updates/sec

Title: US-09-825-414-66

Perfect score: 2407
Sequence: 1 MHINQSAQQPPGVAMESFRT.....EEGGISAFSRSETPFQLRL 487

Scoring table: BLOSUM62

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 08

Listing first 45 summaries

Database : PIR_76:†

1: $\pi_{1r1} =$

2: $\pi_1 r_2 =$

4: pir4:4

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	§					
No.	Score	Query Match	Length	DB	ID	Description
1	149.5	6.2	2155	2	AD2742	conserved hypothet
2	149.5	6.2	2155	2	C97523	hypothetical prote
3	149	6.2	989	2	AE3045	ice nucleation pro
4	149	6.2	1009	2	G98240	hypothetical prote
5	143.5	6.0	2055	2	T31110	extracellular matr
6	141	5.9	536	2	D84325	HtrI7 transducer f
7	140.5	5.8	2186	2	H69960	hypothetical prote
8	140	5.8	2285	2	T12796	probable transglyc
9	138	5.7	1122	2	G64887	probable tail fibe
10	138	5.7	1731	2	AB3045	ice nucleation pro
11	138	5.7	1731	2	B98241	hypothetical prote
12	135	5.6	971	2	B50835	probable tail fibe

13	135	5.6	973	2	C85693	probable membrane
14	134	5.6	2535	2	AC0304	probable hemolysin
15	132.5	5.5	654	2	C87587	hypothetical prote
16	132.5	5.5	1104	2	S59310	probable membrane
17	131.5	5.5	1214	2	G97419	streptococcal hema
18	131.5	5.5	1248	2	AH2637	conserved hypothet
19	131	5.4	409	2	S70548	sapC protein - Sal
20	130	5.4	1026	2	A48995	paracrystalline su
21	130	5.4	1073	2	C87374	S-layer protein Rs
22	129	5.4	436	2	C81655	conserved hypothet
23	129	5.4	637	2	C87322	methyl-accepting c
24	128.5	5.3	2388	2	JE0271	beta spectrin, bet
25	127	5.3	397	2	I39579	nccb protein - Alc
26	127	5.3	1010	2	AH2553	hypothetical prote
27	126.5	5.3	1131	2	T41144	hypothetical serin
28	126.5	5.3	1630	2	A53577	ascites sialoglyco
29	126.5	5.3	2271	2	F90073	hypothetical prote
30	125.5	5.2	902	2	H87323	hypothetical prote
31	125.5	5.2	1225	2	F75605	minor tail protein
32	125	5.2	409	2	S70215	slpC protein - Sal
33	125	5.2	409	2	AH0850	pathogenicity isla
34	125	5.2	1329	2	T29074	hypothetical prote
35	124	5.2	1128	1	T08322	plasmid replicatio
36	124	5.2	2481	2	D90011	FntB protein limpo
37	123.5	5.1	860	2	T14650	tail fiber protein
38	123.5	5.1	978	2	T14968	phage lambda-relat
39	123.5	5.1	1147	2	T35781	hypothetical prote
40	123.5	5.1	2232	2	T34434	hypothetical prote
41	123	5.1	544	2	JC6063	chaperonin groEL -
42	123	5.1	544	2	B83720	class I heat-shock
43	122.5	5.1	789	2	E84236	Htr6 transducer li
44	122	5.1	6713	2	B89921	hypothetical prote
45	121.5	5.0	518	2	A55864	flagellin - Woline

ALIGNMENTS

RESULT 1
AD2742
conserved hypothetical protein Atu1348 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AD2742
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.; Kitajima, J.P.; Okura, V.K.; Almeida Jr., N.F.; Zhou, Y.; Bovee Sr., D.; Chapman, P.; Clendenning, J.; Deatherage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayvin, T.; Levy, R.; Li, M.; McClelland, E.; Palmeri, A.; Raymond, C.; Rouse, G.; Saenphimmachak, C.; Wu, Z.; Gordon, D.; Eisen, J.A.; Paulsen, I.; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, B.; Liao, L.; Kim, S.; Hendrick, C.; Zhao, Z.; Dolan, M.; Tingey, S.V.; Tomb, J.; Gordon, M.P.; Olson, M.V.; Nester, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AD2742
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2155 <KUR>
A;Cross-references: GB:AE006688; P1DN:AAL42354.1; P1D:g17739760; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu1348
A;Map position: circular chromosome

Query Match 6.2%; Score 149.5; DB 2; Length 2155;
Best Local Similarity 20.7%; Pred. No. 0.35;
Matches 96; Conservative 65; Mismatches 176; Indels 127; Gaps 16;

QY	69	DALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAF-----ENTP	124
Db	609	DAFTNSHAKIDTVLAERSNALF-----GALSASQDRFDEALASRSIAITGSVSGTA	659
QY	125	FAAASVLQWQPAINK-----GDWLATPLKPLTPLISGALSGAMDQVGTQMDRARGDLH	179
Db	660	EHLAAMLDRRAAINSVADVERRLLETETRAAAITGAVSGIEDRISDTLESRTAA---716	
QY	180	YLISTSPDKLDAMAVSVKRHSPALGRQVDMGIAVQTFSAINVRTVLAPALASR-----234	
Db	717	-----LHDVSGAESRLADTLD-----GRTAALSSAISGVEERLADTMDSRTLSD	762
QY	235	-----PSVQGANVDFGVSTAGGLVAN-----AGFGDMLSVQSR--DQLRG	272
Db	763	MTFANVEERLSETLDNRTSALTGIVASAEKTAGALDSRTATFGDVAGAETRIETLDG	822
QY	273	-----GAFVLGMKDKEPKAALSEETMLDAYKAIKSASYSGA-----309	
Db	823	RTAALNAVVSGAER-----IADALDSRTMALDMTFSGAEERKIAEALDRTAALGEL	874
QY	310	ALNAGKRMAGLPLDVATDGLKAVRSIVSATSLTKNGLALAGYAVSKIQMATKNTDS	369
Db	875	VASAEETRIAG-ALDSRTDSLKT-----VSGAEEERT-DVLDS	910
QY	370	ATKAASQLSNLVGSVGFAGWTTAGLATDPAVKKAESFIQDKVKTASTTSYVADQTV	429
Db	911	RTMALDMSRSGVEEKITDILDGRTA-----ALKSAVAGVEDRIAGALDSRTAALSG---961	
QY	430	KLAKTVKMSGEAIISSTGASLRSTVNNLRHRSAPeADIEEGGIS	473
Db	962	-IYSGAEERIAEALDSRTALDMTISGVEERIAEAMDARAASSLS	1004

RESULT 2
C97523
hypothetical protein AGR_C_2490 [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C;Accession: C97523
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, B.S.; Cao, Y.; Askenazi, M.; Halling, C.; Mullin, L.; Houmiel, K.;

RESULT 4
G98240
hypothetical protein AGR_L_1758 (imported) - Agrobacterium tumefaciens (strain C58, Cereon)
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C:Accession: G98240
R:Goodner, B.; Hinkley, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, B.S.; Cao, Y.; Askenazi, M.; Halling, C.; Mullin, L.; Houmief, K.; Gordon, J.; Vandin, M.; Iarchouk, O.; Epp, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Flanagan, C.; Crowell, C.; Gusion, J.; Lomo, C.; Sear, C.; Strub, G.; Clelo, C.; Slater, S.
Science 294, 2323-2328, 2001

G98240

A;Map position: linear chromosome

Matches 127; Conservative 77; Mismatches 228; Indels 136; Gaps 25

• •
—

—
—

—

—
—
• •

• •
—
—

—

—

• •

Db 617 KAITTKALGELDVTKLSTGNIAALTKAQ 644

T31110

A;Gene: emb

Db 786 AKYPEALGH--VRQADAKKQAIKDANLTAEEQADALRQVDAQTAAEAAINQNHTNA 842

[illegible]

A;Gene: htrI7
C;Superfamily: Halobacterium salinarum transducer protein htrI

Query Match 5.9%; Score 141; DB 2; Length 536;
Best Local Similarity 23.2%; Pred. No. 0.17;
Matches 119; Conservative 78; Mismatches 197; Indels 120; Gaps 23;

QY 12 GVAMESFRTASDASLSSSVRSVSTSCRDLOAITDYLKHHFAHRFSVIGSPDERDA 71
 |::|| | : : : : : || | | | |
DB 63 GPWMDAFST----SLLAVLGVVAISSTGTISIVED-LKTSNARHR-----RRDAEQAK 113

QY 72 LAHNEQIDALVETRANRLYEGETPATIAETFAKEEFDRLATTASSAENTPEPAASVL 131
 :|| | : | : | : : | : | : : : : || |
DB 114 RRADEQRAQAQAEAKAEAQDAREAEITAAQLQERRAERDAQLSKASEYSSVMEKAAS-- 171

QY 132 QYNQPAINKGDMLATPLKPLPLI -SGALSAMDQVG-----TWMDRBR----- 175
DB 172 -----GD-----LTVRANSSSDSDATEVGAFNEMI AAFEEMVARTREFAGD 214

QY 176 -----GDLHYLSTPDKLHDAMAVSVKRHPALGRQVYDMGIAVQTFSALNVRTVLAPA 230
DB 215 VADNSRDIHASVDEIEAASEQVAESVQEISAQTDRE-----HTRLADA 257

QY 231 LASRPSVOGAVDFGVSTAGGLVANAGFGDRMLSVQSQRDLRGCAFVLMKKDKPEKPAALSE 290
DB 258 TDELQSLSGAVE-EVAASASQVAD-----VSDQAADRGRGS-----DDAAAAVTQ 302

QY 291 ETTWLDAYKAIKS-ASYSGAALNAGRWAGLPDVATD--GLKAVRSIVSATSLTKNGLA 347
DB 303 MTDLQSETAEVNSEANALAEAQEIIRVVLDLINDI-TDQTHLLALNASLEAARAQDDGDG 361

QY 348 LAGGYAGVSKLQNKATKNITD-----SATKAVSQLSNLVGSVGFAGMTTAGLA 397
DB 362 FAYVAAEVEKELAE-ETNEATDEIETLVNLLSSTQSVVDSEQMAYTS--SGETVENA 418

QY 398 -----TDPAYKCAASFIODKVKSTASTSYVADQIVKLAKTVKDMSGEAIST 446
DB 419 LTAAEEIGHVTD--VNEAVQOIDDATDAQAAS-----ADTTVMIEIEDIS-EQNSSE 470

QY 447 GASLRSTVNNLRHRSAPAEADIEGGISAFSRSET 480
DB 471 AASVSAAA---QQQTASLSPTVDGVTRLVLRADT 501

RESULT 7
HB9960
hypothetical protein SA1577 [imported] - Staphylococcus aureus (strain N315)
CISpecies: Staphylococcus aureus
CDate: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
CAccession: HB9960
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui,
L.; Oguchi, A.; Aoki, K.; Nagai, Y.; Lián, J.; Ito, T.; Kanamori, M.; Matsumaru,
H.; Matsuyama, A.; Murakami, H.; Hosoyama, A.; Mizutani-Ui, Y.; Kobayashi, N.;
Swano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; Hirakawa, H.; Kuhara, S.; Goto,
S.; Yabuzaki, J.; Kanehisa, M.; Yamashita, A.; Oshima, K.; Furuya, K.; Yoshino,
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hirataetsu, K.
Lancet 357, 1225-1240, 2001

T12796
probable transglycosylase - *Bacillus subtilis* phage SPBc2
C/Species: *Bacillus subtilis* phage SPBc2
C/Date: 13-Aug-1999 #sequence _revision 13-Aug-1999 #text_change 15-Oct-1999
C/Accession: T12796; A65911
R/Lazarevic, V.; Duesterhoeft, A.; Soldo, B.; Hilbert, H.; Maue, C.; Karamata,
D.
submitted to the EMBL Data Library, August 1997
A/Description: The complete nucleotide sequence of the *Bacillus subtilis*
SPBc2 prophage.

submitted to the EMBL Data Library, August 1997
A:Description: The complete nucleotide sequence of the *Bacillus subtilis* SPbetac2 prophage.
A:Reference number: Z17583
A:Accession: T12796
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA

A;Cross-references: EMBL:AF020713; NID:g3025478; PID:g3025510; PIDN:AA033005.1
R;Kunst, F.; Ogawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, G.; Galleron, N.; Gim, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, A.; Giuseppe, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maeneel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, A.M.; Paek, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott, S.; Riegler, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.-J.; Serro, P.; Shin, B.-S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Tempstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannieu, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A;Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: A69911
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2285 <KUN>
A;Cross-references: GB:Z29115; GB:AL009126; NID:g2634478; PIDN:CAB14053.1;

A;Experimental source: strain 168
C;Genetics:

```

R.;Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Butland, V.;
Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor,
J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: G64887
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1122 <BLAT>
A;Cross-references: GB:AE000234; GB:U00096; NID:g1787633; PIDN:AAC74454.1;
PID:g1787636; UWGP:b1372
A;Experimental source: strain K-12, substrain MG1655
R;Alba, H.; Baba, T.; Fujita, K.; Hayashi, K.; Inada, T.; Isono, K.; Itoh, T.;
Kasai, H.; Kashimoto, K.; Kimura, S.; Kitakawa, M.; Kitagawa, M.; Makino, K.;
Miki, T.; Mizobuchi, K.; Mori, H.; Mori, T.; Motomura, K.; Nakade, S.; Nakamura
Y.; Nashimoto, H.; Nishio, Y.; Oshima, T.; Saito, N.; Sampei, G.; Seki, Y.;
Sivasubdaram, S.; Tagami, H.; Takeda, J.; Takemoto, K.; Takeuchi, Y.; Wada, C.;
Yamamoto, Y.; Horiuchi, T.
DNA Res. 3, 363-377, 1996
A;Title: A 570-kb DNA sequence of the Escherichia coli K-12 genome correspondin
to the 28.0-40.1 min region on the linkage map.
A;Reference number: Z16603; MUID:97251357; PMID:9097039
A;Accession: T09189
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 3-1122 <AIB>
A;Cross-references: GB:AE000234; GB:U00096; NID:g1787633; PIDN:AAC74454.1;
PID:g1787636; UWGP:b1372

Query Match          5.7%; Score 138; DB 2; Length 1122;
Best Local Similarity 21.1%; Pred. No. 0.76;
Matches 106; Conservative 65; Mismatches 221; Indels 110; Gaps 17;

QY      3  INOSAQQPPGVAMESEFRTASDASLASSVRSVSTTSCRDLQAITDYLKHVFAHRSVI 62
      : : : : : : : : : : : : : : : : : :
Db      112 VEEVARNASVAQNTAAKKKSADASTAREAAHTAA----- 148

QY      63 GSPDERDAALAHNEQIDALVETRANRLYSEG--ETPATIA-ETFAKAKEFDRLATTASSA 119
      | : | : | : : : : : | | | | : : | : | : |
Db      149 ---DAADSARAASSTSAQMASSAQSASSAGTASTKATEASKSAALAAESSKSAATSA 205

QY      120 FENTPFAAASVLQVQPAINKGDMLATPLKPLTPLISGALSGAMDQVGTKKMDRARGDLH 179
      : : : : : : : : : : : : : : : : : :
Db      206 AKTSETNNAASLQS-----AATASASTATTKASEAATSARDAASK--EAKKSET 253

QY      180 YLSTSPD-----KLHDAMAVSVKRHSPALGROY-VDMGIAVOTFSALNVRT 225
      : : | : : : : : : : : : : : : : : : :
Db      254 NASSSASSAASATTAAGNSAKAAKTSETNARSETTAAGOSAAAGSKTTAAASSASAST 313

QY      226 VLAEPALASRPSVGAVDFGVSTAGLIVANAGFEDRLSVQSRDQLRGAGFYLGAKDKPEK 285
      | | : : : : : : | | | : : : | : |
Db      314 SAGQASASATTAAGKSAESMASSASTATTKAGEATEQASAAARS-----ASAAKTSETN 366

QY      286 AALSEETDWLDAYKAIKSASYSGAALINAGKRMAGLPLDVATDGLKAVRSIVSATSLTNG 345
      | | : : : : : : | | | : : : | : | : : :
Db      367 AKASETS--AESKTTAAASSASAASSASS--ASASKDEATRQASAAKS--SATTAATKA 420

```

QY 346 LALAGYAGVSKLQKMATKNITDSATKAVSOLSNLVGSVGFAGWTTAGLATDPAVKKA 405
11 111 11 1 : :
Db 421 TEAG-----SATAAQSSTA-----ESAATRAETAAKRA 451
QY 406 ESFIQDKVKSTASTSYVADQYVKLAKTVKDMGSAISSTGASLRSTVNNLHRSAPE- 464
1 11 1 : : 1 : : 1 : : 1 : : 1 : :
Db 452 EDIASAVALEDASTTKGI---VQLSSATNSTS-ETLAATPKAVKSAYDNAEKRLQKQ 506
QY 465 --ADIEEG-----ISAFSRSE 479
11 1 : 1 1 : : :
Db 507 NGADIPDKGCFUNNINAVSKTD 528

RESULT 10

AB3045
ice nucleation protein homolog [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AB3045
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.; Kitajima, J.P.; Okura, V.K.; Almeida Jr., N.F.; Zhou, Y.; Bovee Sr., D.; Chapman, P.; Clendenning, J.; Deatherage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayvin, T.; Levy, R.; Li, M.; McClelland, E.; Palmeri, A.; Raymond, C.; Rouse, G.; Saenphimmachak, C.; Wu, Z.; Gordon, D.; Eisen, J.A.; Paulsen, I.; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, B.; Liao, L.; Kim, S.; Hendrick, C.; Zhao, Z.; Dolan, M.; Tingey, S.V.; Tomb, J.; Gordon, M.P.; Olson, M.V.; Nester, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AB3045
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1731 <KUR>
A:Cross-references: GB:AE008689; PIDN:AAL44776.1; PID:g17742414; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: ina
A:Map position: linear chromosome

Query Match 5.7%; Score 138; DB 2; Length 1731;
Best Local Similarity 21.4%; Pred. No. 1.4;
Matches 116; Conservative 76; Mismatches 197; Indels 152; Gaps 23;
QY 1 MHINQSAQQPPGVAMESFRTASDASLASSVSSTSCRDLQAITDYLKHVFAAHRFS 60
111 : : : 1 1 : : : 1 : : : 1 : : : 1 : : :
Db 746 MHIAELGTED---LAKFTTKDIAAIISSSAIAGLSA-----EAIASLTTAQIAALNTQS 795
QY 61 VIGSPDERDAALAHNEQIDALIVETRAANRL-----YSEGETPATIAETFA- 104
: : : 1 : : : 1 : : : 1 : : : 1 : : :
Db 796 INALSTAQIAALT-TAQVEALFTSQVNALTSKQIAALSTDIATFTKIDIAAINSDAIAIG 854
QY 105 -KAEKDRLATTAASSAFENTPFAAASVLQYMQPAINKGDWLA-T-PLKPLTPLISGALSGA 162

Db 855 LSAETIASLATQIAPALNVRSIAALSTVQIVA-----LTTAQVEALTTVQVGLSST 906
11 111 : 1 1 1 : : : 1 : : : 111
QY 163 MDQVGTKNMDRARGDLHYLSTSPDKLHDAMAVSVKRHSPALGRQV----- 208
Db 907 -----QLAVLST-----DDIATFSTRDMAALGSSAIAGLSKDTVASLTTA 946
QY 209 DMG-----IANTQFSALNVRTYLPALASRP-----SVQGANDFGVSTA 248
1 : 1 111 : 1 1 : : 1 : : : 1 : : :
Db 947 QIGALSMAGISGLSTGQIAALTGDQINVLTNTQIAALTSKVAAAFDVSIDITLSTQIAA 1006
QY 249 GGLVANAGFDRMISVQSRDQLRGAFVLGKKDKPEPKALSEE-----TDLDAVKAIK 302
11 : : 1 11 : 1 : : : 111 : 11 : :
Db 1007 LSAAGAAGLTTDQIAALSTDOV--GAMTSGQ-----IAALSAKQIAALGTDIDIAFTSTGD 1059
QY 303 SASVSGAALNAGKRMAGLPIDVATDGLKAVRSIVSATSILTNGLALAGY----- 352
1 : 1 1 : : 111 1 : : : : 1 : : :
Db 1060 IAAISSNA-----VAGLSRDTVASLTTAQIAALSSAGISGLGTQIAGLTSQVNVLTN 1113
QY 353 AGVSKL--QKMATKNITDSATKAAVSOLSNLVGSVGFAGWTTAGLATDP----AVKKA 406
1 : 1 1 : : : 111 : : 1 : : : 111 11 : : :
Db 1114 AQISALTSKVAAALEVTDIASLSA-AQIA-AIGAAGV-----AGLTTDQIAALSTSQVE 1165
QY 407 SFIQDKVKSTASTSYVADQTVKL-----AKTVKDMGSAISSTGA 448
: : : 1 : : : : : : : : : : : : : : :
Db 1166 ALTSAQIAALNSKQIAALSADDLAIFTTAEAMAIISGALSGLPASTIASLTTAQIAALGA 1225
QY 449 S 449
Db 1226 A 1226

RESULT 11

B98241
hypothetical protein AGR_L_1764 [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C:Accession: B98241
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, B.S.; Cao, Y.; Askenazi, M.; Halling, C.; Mullin, L.; Houmlel, K.; Gordon, J.; Vaudin, M.; Iartchouk, O.; Epp, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Flanagan, C.; Crowell, C.; Gursen, J.; Lomo, C.; Sear, C.; Strub, G.; Cielo, C.; Slater, S.
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens C58.
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: B98241
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1731 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK89452.1; PID:g15159316; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L_1764
A:Map position: linear chromosome

C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: C85693
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, G.F.; Evans, P.S.; Gregor, Y.; Kirkpatrick, H.A.; Posfai, G.; Hackett, J.; Klink, S.; Boutin, A.; Snao, Y.; Miller, L.; Grobeck, E.J.; Davis, N.W.; Iim, A.; Dimalanta, E.; Potamouzis, K.; Apodaca, J.; Anantharaman, T.S.; Lin, J.; Yen, G.; Schwartz, D.C.; Welch, R.A.; Blattner, F.R.
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: C85693
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-973 <STO>
A;Cross-references: GB:AE005174; NID:g12514847; PIDN:AA656007.1; GSPDB:GN00145; UWGP:21918
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: 21918

Query Match 5.6%; Score 135; DB 2; Length 973;
Best Local Similarity 22.0%; Pred. No. 0.96;
Matches 116; Conservative 57; Mismatches 238; Indels 116; Gaps 20;

QY 3 INQSAQPPGVAMESFRTPASDASLASSSVRSVSTSCRDLQAITDYLKHAFVFAHRFSVI 62
: : | : | : | | | | : |
DB 112 VEEVARNASAVAQNTAAAKKSASDASTSAREEAT----- 145

QY 63 GSPDERDAALAHNEQIDALIVETRANRLYSEG--ETPATIETFAKAEKFDRLATTASSAF 120
: | | | : : : | | | | | : : | :
DB 146 HATDAADSDARAASTSAGQAASSAQSSASAGTASTKATEASKSAAAAESSKSAATSGA 205

QY 121 ENTFFAASVLYQWQPAINKGWLATPLKPLPLISGALSGAMDQVGTMMMDRAGDLHY 180
| | | | | | | | | | | | | | : : :
DB 206 AKTSETNAAVSQ--QSA-----ATSASTATTKASEAASARDASASKEAAS--SETSA 255

QY 181 LSTSPDKLHDAAV--SVK-----RHSPALGRQVNDMGIAVQTFESAL--NVVRTVL 227
: : | | | | | : : | : | : | : |
DB 256 ASSASSAASSATPAAGNSAKAAKTSETNAKSSETPAEQSASAAAAGSKTAAALSASAASTSA 315

QY 228 APALASRPSVQGVDFGVSTAGLVANAGFGDRMLSVQSRDQLRGAFVLGMKDKEPKAA 287
| | : : : : | : | : : : : | : |
DB 316 GQASASATPAAGKSAESAASSASTATTKAGEATEQASA-----AASSASAAKTSETNAK 368

QY 288 LSEETDWLDAYKAIKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLSVATSLTKNGLA 347
| | : : | : | | | | : | | | : | : | :
DB 369 ASETS--AESSKTAAASSSAASSAASS--ASASKDEATROASAAS--SATTASTKATE 422

QY 348 LAGGYAGVSKLQ-----KQATKNITD-----SATKAAVSOLSNLVGSVGV 387
| | : : : : : | | | : | | : | : | :
DB 423 AAGSATAAQSKSTAESAATRAETRAKRAEDIASAVALEDASTTKGIVQLSASTNS--- 479

QY 388 FAGWTTAGLATDP-AVKKAESFIQDKVKSTASSTTSYVADQTVKLAKTVKDMSGEALISST 446
| : | | | | | | | | | : | : | : | : | :
DB 480 ----TSESLAATPKAVKAVYELANGK-----YTAQDATTAQKGIVQLS-NATNST 524

QY 447 GASLRSTVNNLR-----HRSAPREADIEEGCI-----SAFSRSET 480
| : | : : : : | : | : | : | : | : | : | :
DB 525 SEMLAATPKSVKAAVYDLANGKYTAQDATTAQKGIVQLSSATNSASET 571

RESULT 14
AC0304
probable hemolysin YPO2490 [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C;Accession: AC0304
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tiltball, R.W.; Holden, M.T.G.; Prentice, M.B.; Sebahia, M.; James, K.D.; Churcher, C.; Mungall, K.L.; Barker, S.; Basham, D.; Bentley, S.D.; Brooks, K.; Cerdeno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Feltham, T.; Hamlin, N.; Holroyd, S.; Jagers, K.; Leather, S.; Karlyshev, A.V.; Moule, S.; Oyston, P.C.F.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell, B.G.
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AC0304
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2535 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC91295.1; PID:g15980484; GSPDB:GN00175
C;Genetics:
A;Gene: YPO2490

Query Match 5.6%; Score 134; DB 2; Length 2535;
Best Local Similarity 23.0%; Pred. No. 4.2;
Matches 84; Conservative 66; Mismatches 168; Indels 48; Gaps 16;

QY 126 AAASVLYQWQPAINKGWLATPLKPLPLISGALSGAMDQVGTMMMDRAGDLHYLSTSP 185
| : : : : : | | | : | : | : : | : | :
DB 1383 ATSEHI RHIGSEVNVGALTNVNDLTL-AVGANINAAITLEVQAQNIS-----LSAAT 1433

QY 186 DKLHDAAVSVKRRHSPALGRQVNDMGIAVQTFESALNVVRTVLAPALASRPSVQGVDFGV 245
| | | | | | : : : : : | : : | : | : |
DB 1434 DSIHVTGESSKRHTSSV--NLXDETLLGSQLNATGDNILQAAQDITLIRAS-----AV 1484

QY 246 STAGGLVANAGFGDRMLSVQS--RDQLRGAFVLYGMKDKEPKALSEETDWLDAYKAIKS 303
| | | | | | : | : | : | : | : | : | :
DB 1485 QTDGALLTLAAG-GDVLTLTQTEQHDEQRNHT---GLSKGIASSSTLRTEDSLQTLAVGS 1540

QY 304 ASYSGAALNAGKRMAGLPLDVATD---GLKAVRSLSVATS-----LTKNGLALA 349
: | : : | : | : : | | : : | : : | :
DB 1541 MLSAGSIDVSGKNIAVMGSNVVADQDISLRAQENITVGTAAQOSESESHLFEQKSGLMST 1600

QY 350 GGYAGVSKLQKMATKNITDSATKAAYSQLSNLVGSVGFAGWTTAGLATDPAVKKAESFI 409
| | | : : : | : | : : | : | : | : | :
DB 1601 GG-IGVT-VGSSSTK-MTDSG--QSISVSGTIVGSV---LGNVSMTAGEDLRVQGAEVLA 1652

QY 410 QDKVKSTASSTTSYVADQTVKLAKTVKD-MSGEALISSTGASLRSTVNNLRHRSAPREADIE 468
: | : : : : | : : : | : : | : : | : | :
DB 1653 GKDINLTGKNVSIILAEENQLTQSHTVBQKSGILTLALSGA-VGSAVNTAVTTAKAASEES 1711

QY 469 EGGISA 474
| : |
Db 1712 SGRIGA 1717

RESULT 15
C87587

hypothetical protein CC2730 [imported] - Caulobacter crescentus

C;Species: Caulobacter crescentus

C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C;Accession: C87587

R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.F.; Alley, M.; Ohta, N.; Maddock, J.R.; Potocka, I.; Nelson, W.C.; Newton, A.; Stephens, C.; Phadke, N.D.; Ely, B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolonay, J.F.; Smit, J.; Craven, M.; Khouri, H.; Shetty, J.; Berry, K.; Utterback, T.; Tran, K.; Wolf, A.; Vamathevan, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A;Title: Complete Genome Sequence of Caulobacter crescentus.

A;Reference number: A87249; MUID:21173698; PMID:11259647

A;Accession: C87587

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-654 <STO>

A;Cross-references: GB:AE005673; NID:g13424321; PIDN:AAK24695.1; GSPDB:GN00148

C;Genetics:

A;Gene: CC2730

Query Match 5.5%; Score 132.5; DB 2; Length 654;

Best Local Similarity 24.4%; Pred. No. 0.8;

Matches 116; Conservative 67; Mismatches 162; Indels 131; Gaps 28;

QY 61 VIGSPDERDAALAHNEQIDALVETRANRLYSEGETPATIAET---FAK--AEKFDRLAT 114
| : | | : : | | | : | : | | | : :
Db 3 VVGS-----QNTALSSISPTRIAL-----SAATISEAPASEFAKLTAAQVAALSI 48

QY 115 TASSAFENTPFAAASVLIQYMOPAINKGDWLATPLKPLT----PLISGALSGANDQVGTKM 170
| | | : | | | | : : | | | : | : | : | :
Db 49 TALSALTSSEQFAAFDVRQ-----VRALTSAQULPRLSAALISFSPKQLAT-- 91

QY 171 MDRARGDLHYLSTSPDKLHDMAVSVKRHSPALGR-QVDMGIAVQT-FSALNV----- 222
| | | : : | : | | | : : | : | : | : | :
Db 92 -----LSQAQTFQALQHTQVAALSI-----TALGEMRTQQLAVLSATFAALSIDQVOSL 140

QY 223 ----VRTVLAPALASRPSVQAVDFGVSTAAGLVA--NAGFGDRMLSVQSRDQLRGAFV 276
| : | : | : | : | : | : | : | : | : | : | : | :
Db 141 SLAQVRSLSAPQVRALPADI--EFSLEQVGAITAIQLAVLSPQVSALSSSEQ----- 192

QY 277 LGMKDEPKAALSEET-----DWLDAYKAIKSASYSGAALNA--GKRMAGL 320
: | : : | | | | : : | : | : | : | : | : | :
Db 193 IGALSRTOFSALSAATVRALSPQTLATTPADHWRRAATIAQMSSLSGDQLKALGETRLSSL 252

QY 321 PLDVATDGLKAVRSIVSATSLTKNGLALAG-----YAGVSKLQKMATKNITDSATKAAY 375
| : | | | | : : | | | : : | : | : | : | :
Db 253 -----SASQVRAL-SATEVGRAGPALDLALTPTQLAGLGAIH-TAALTVTDIAGLSA- 302

QY 376 SOLSNLVGSVGVFAGWTIAGLATDPAVKKAESFIQDKVKSTASTTSYVADQTVKL-AKT 434
| : : | | : | : | : | : | : | : | : | : | :
Db 303 SQIAAL--DPAAVAGLSATGLSA-----LSASQVAALASASQVARKLRANQ 344

QY 435 VKDMSGEALISS-TGASLRS-TVNNLRH-RSAP EADIEGGISAFSRSETPPQQLRRL 487
: : : | | | | : : | : | : | : | : | : | : | :
Db 345 LQGLNGEDFSEFTPAQLKAFTTAAQIRELP SARLAELSDDEDfGEF----TATQIRAL 396

Search completed: January 6, 2004, 09:46:19
Job time : 24 secs

OM protein - protein search, using sw model

Run on: January 6, 2004, 09:37:23 : Search time 18 Seconds
(without alignments)
1272.334 Million cell updates/sec

Title: US-09-825-414-66
Perfect score: 2407
Sequence: 1 MHINQSAQQPPGVAMESFRT.....EEGGISAFSRSEMPQLRL 487

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	138	5.7	1120	1	STFR_ECOLI P76072 escherichia
2	137	5.7	540	1	CH60_LACHE 068324 lactobacilli
3	130	5.4	500	1	FLJB_SALAE P52615 salmonella
4	130	5.4	1025	1	SLAP_CAUCR P35828 caulobacter
5	129	5.4	436	1	Y868_CHIMU Q9pjg1 chlamydia m
6	128.5	5.3	2388	1	SPCP_RAT Q9qwm8 rattus norv
7	128	5.3	546	1	CH60_LACAC Q93g07 lactobacilli
8	127	5.3	397	1	NCCB_ALCXX Q44585 alcaligenes
9	123	5.1	544	1	CH60_BACHD O50305 bacillus ha
10	121.5	5.0	1068	1	HIPR_MOUSE Q91ky5 mus musculu
11	120.5	5.0	1140	1	YMR6_YEAST Q04893 saccharomyc
12	120	5.0	1306	1	MSB2_YEAST P32334 saccharomyc
13	119.5	5.0	1608	1	HLYA_SERMA P15320 serratia ma
14	119	4.9	657	1	HS7F_CAEEL P11141 caenorhadi
15	118.5	4.9	760	1	METE_MYCLE O05564 mycobacteri
16	118.5	4.9	880	1	LYTD_BACSU P39848 bacillus su

17	117	4.9	583	1	YP65_MYCTU Q50733 mycobacteri
18	117	4.9	682	1	GR78_YEAST P16474 saccharomyc
19	116.5	4.8	570	1	FLIF_RHOSH Q53151 rhodobacter
20	116	4.8	763	1	HTR2_HALNI Q9hp81 halobacteri
21	116	4.8	764	1	HTR2_HALSA P71410 halobacteri
22	116	4.8	1577	1	HLXA_PROMI P16466 proteus mir
23	116	4.8	3591	1	PHAB_BORPE P12255 bordetella
24	115.5	4.8	401	1	YOPB_YERPS Q06114 yersinia ps
25	115.5	4.8	1065	1	SED4_YEAST P25365 saccharomyc
26	115	4.8	679	1	GR78_KULIA P22010 kluyveromyc
27	115	4.8	1239	1	V120_EBV P56963 campylobact
28	113.5	4.7	571	1	FLAI_CAMJE Q9y490 homo sapien
29	113	4.7	2541	1	TLN1_HUMAN P52616 salmonella
30	112.5	4.7	505	1	FLJB_SALTY P35658 homo sapien
31	112.5	4.7	2090	1	N214_HUMAN O84583 chlamydia t
32	112	4.7	439	1	Y579_CHLTR P06175 salmonella
33	112	4.7	492	1	FLIC_SALRU P55197 homo sapien
34	112	4.7	1027	1	AF10_HUMAN Q9K2u4 propionibac
35	111.5	4.6	544	1	CH60_PROAC Q9xau7 alteromonas
36	111.5	4.6	547	1	CH60_ALTHA P07017 escherichia
37	111.5	4.6	553	1	MCP2_ECOLI P47033 saccharomyc
38	111	4.6	881	1	PRY3_YEAST P47632 mycoplasma
39	110.5	4.6	543	1	CH60_MYGE O13368 candida alb
40	110.5	4.6	1419	1	ALAI_CANAL Q48318 halobacteri
41	110	4.6	545	1	HTR5_HALNI Q93gt8 colwellia m
42	110	4.6	549	1	CH60_COLMA P49051 bacillus an
43	110	4.6	814	1	SLAI_MOUSE Q91zu6 mus musculu
44	110	4.6	7389	1	BPAL_MOUSE P40317 saccharomyc
45	109.5	4.5	901	1	SOK1_YEAST

ALIGNMENTS

RESULT 1	STFR_ECOLI	STANDARD;	PRT; 1120 AA.
ID	STFR_ECOLI AC	P76072; P77560;	
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Side tail fiber protein homolog from lambdaoid prophage Rac.		
GN	STFR OR B1372.		
OS	Escherichia coli.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Escherichia.		
OX	NCBI_TaxID=562;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=K12 / MG1655;		
RX	MEDLINE=97426617; PubMed=9278503;		
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,		
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,		
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,		
RA	Mau B., Shao Y.;		
RT	"The complete genome sequence of Escherichia coli K-12.";		
RL	Science 277:1453-1474(1997).		

DR HAMAP; MF 00600; -; 1.
DR InterPro; IPR001844; Chaprinin_Cpn60.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR Pfam; PF00118; cpn60_TCP1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW Chaperone; ATP-binding.
SQ SEQUENCE 540 AA; 57638 MW; 4257DDB45FC7C4B4 CRC64;

Query Match 5.7%; Score 137; DB 1; Length 540;
Best Local Similarity 20.8%; Pred. No. 0.4;
Matches 109; Conservative 72; Mismatches 198; Indels 144; Gaps 24;

QY 21 ASDASLASSSVRSVSTTSCRDQAITDYLKHHVFAHRESVI-----GSPDERDALAHNE 76
| | : : : | | : : : | | : | | : | | : : :
Db 2 AKDIFSENAARSL-----LKGVDKLADIVKTTIGPKGRNVVLEQSYGNPDI TN DGVTI AK 57
QY QID-----ALVETRANRLYSEGETPATIAETFAKAKEFDRLATTAS-----S 118
| : : : | : : : | | | | : : : | :
Db 58 STELKDRYENNGAKLVAAEAQAQKNDIAGDGTATATVL TQA IAREGKNV T AGANPVGI RR 117
QY 119 AFENTPEAAASVLYQMYQPAINKGEMLA--TPLKPLTPLISGALSGAMDQVG----- 167
| | | : : : | : | : : | : : : | | : : :
Db 118 GIEKATKAAMDVDELHKISHVESKQDIANYAAVSSASKI GALIADAMEKVGHDVITIED 177
QY 168 -----TKMDRARGL-HYLSTSPDKLH-----DAMAVSVKRHSPALGR 205
: | | | : | | : | | : : : | | :
Db 178 SRGINTEL SVEGMQFD RGYLSQMYMTDNDKMEADLDNPFYLLITDKKISNIQDILP L L-Q 236
QY 206 QVDMGIAVQTFSAI NVRTVLA PALAS--RPSVQGA VDFGVSTAGGLVANAGFGD-RML 262
: : | | : | : : : | | : : : | | | |
Db 237 EIVQGG-----KSLIIADDDITGELAPTLV LNKIRGT--FN VVA-----VKAPGFGDRRKA 285
QY 263 SVQSRDQLRGAFV---LGMKKEPKKAIS EETMDLDAVKAIKSASVSGAALNAGKRMAG 319
: | | : | | : | | : | | : : :
Db 286 QLDIAALTG GTVITTEDLGELKDTK----- 311
QY 320 LPLDVATDGLKAVRSIVSATSLTRNGLAAGYAG-----VSKLQKMATKNITTSAT 371
| | : : : | | : : | | : : : | :
Db 312 -----IDQLGQARRI-----TVTKDSTTI VGG-AGSKEALDERVDTRKQIEDSTSDPK 360
QY 372 KAAVSQLSNLVGSVGFAGMTTAGLATDPAVKKAESFIQDVKST-ASSTTSYVA---DQ 427
| : : | | | | : : : | : | : | | : : :
Db 361 KKLQERLAKLTGGVAVI---HVGAATETELKERRYRIEDALNSTRAAVDEGVVAGGCTA 416
QY 428 TVKLAKTVKMGSGEAL-SSTGASLRSTVNNLRHRSAP EADIEE 469
| : | | : : | | : : : | | | | |
Db 417 LWNVEKAVREVGETTDEQTGINI-----VLRA LSA PV RQ IAE 454

RESULT 3
FLJB_SALAE STANDARD; PRT; 500 AA.
ID FLJB_SALAE
AC P52615;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Phase-2 flagellin.
GN FLJB.
OS Salmonella abortus-equi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=607;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93139045; PubMed=8423149;
RA Okazaki N., Matsuo S., Saito K., Tomimaga A., Enomoto M.;
RT "Conversion of the Salmonella phase 1 flagellin gene fljC to the
RT phase 2 gene fljB on the Escherichia coli K-12 chromosome.";
RL J. Bacteriol. 175:758-766(1993).

RL [2]
RP SEQUENCE OF 478-501 FROM N.A.
RX MEDLINE=93173100; PubMed=8437573;
RA Hanafusa T., Saitko K., Tomimaga A., Enomoto M.;
RT "Nucleotide sequence and regulated expression of the Salmonella fljA
RT gene encoding a repressor of the phase 1 flagellin gene.";
RL Mol. Genet. 236:260-266(1993).
CC -|- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
CC -|- FORM THE FILAMENTS OF BACTERIAL FLAGELLA.
CC -|- MISCELLANEOUS: INDIVIDUAL SALMONELLA SEROTYPES USUALLY ALTERNATE
CC BETWEEN THE PRODUCTION OF 2 ANTIGENIC FORMS OF FLAGELLA, TERMED
CC PHASE-1 AND PHASE-2, EACH SPECIFIED BY SEPARATE STRUCTURAL GENES.
CC -|- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) announce/
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; D13690; BAA02848.1; -.
DR EMBL; D12510; BAA02072.1; -.
DR PIR; S30923; S30923.
DR InterPro; IPR001029; flagellin_C.
DR InterPro; IPR001492; flagellin_N.
DR Pfam; PF00700; flagellin_C; 1.
DR Pfam; PF00669; flagellin_N; 1.
DR PRINTS; PR00207; FLAGELLIN.
DR ProDom; PD000316; flagellin_C; 2.
KW flagella.
FT INIT MET 0 BY SIMILARITY.
FT CONFLICT 493 493 Q -> L (IN REF. 2).
SQ SEQUENCE 500 AA; 51927 MW; 336D6F8F0EAA56CC CRC64;

Query Match 5.4%; Score 130; DB 1; Length 500;
Best Local Similarity 21.9%; Pred. No. 0.93;
Matches 114; Conservative 71; Mismatches 232; Indels 104; Gaps 24;

QY 2 HINQSAQQPPGVANE-----SFRASDASLASSSVRSVSTTSCRDQAITDYLKHHVFA 55
: : | | | | | | : | | : : : | : : :
Db 16 NINLKS-QSALGTALERLSSGLRINSAKD-DAAQQA IANRFTANI KGLTQASRNANDGISI 73

QY 56 AHRFSYIGSPPERDAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATT 115
|| : | : | : | : : : | || | : : | : : | : | :
Db 74 AH--TTEGALMEINNLL--QVRRELAVQSANSTNSQSDLDSIQAEITQRLNEIDRV--S 126
QY 116 ASSAFENTPFAAASVLQYMQPAINKGWDLATPLKPLTPLISGA-----LSGAMDVGTKM 171
: | : | : | : | : | : | : | : | : | : | : | :
Db 127 GQTQFNQVKVLAQNDNTLTIQVANGDEGTIDIDLQINSQTLGLDSLNVQKADVATDVI 186
QY 172 DRARGDLHYLSTSPDKLHDAMAVSVKRHSPALGRQV--DMGIAVQTFESALNVRTVLAP 229
| : | : | : | : | : | : | : | : | : | : | : | :
Db 187 SSTYSDDGTQALTAP-----TATDIK---AALGNPTVTGDTLTAAVSVFKDGKYATVSGY 237
QY 230 ALASRPSVQGAVDGFGVSTAGGLVANAGFGDRMLSVQSRDQLRGAFYLGAKDKEPKAALS 289
| : | : | : | : | : | : | : | : | : | : | : | :
Db 238 TDAGDTAKNGKYEVTVDSATGAVS--FG---ATPTKSTVTGDTAVTKVQVNAVPAA-- 288
QY 290 EETDWLDAYKAIKSASYSGALINAGKRMAGLPLDVATDGLKAVRSIVSATSLTKNGIALA 349
|| | | | | : | : | : | : | : | : | : | : | :
Db 289 -----DA--ATKKALQDGVSSADASAA-----TIVKMSYTDKNGKTIE 325
QY 350 GGYA-----GVSRLQKMATKNITDSATKAAVSQLSNLVGSVGFAGWT 393
||| | : | : | : | : | : | : | : | : | : | :
Db 326 GGYALKAGDKXYAADYDEATGAIR-AKTSYTAADGTTKTAANQLGGVDGKTEVV---TI 381
QY 394 AGLATDPAVKKAESF-IQDKVKSTASTSTSYVADQ-TVKLAK-----TVKDMSGEA 442
| : : : | : | : | : | : | : | : | : | : | :
Db 382 DGKTYNASKAAGHDFKQPELAELAAKTENTPLQKIDALLAQVDALRSDLGAVQNRENSA 441
QY 443 ISSTGASLIRSTVNNLHRHSAPEADIIEGC---ISAFSRSE 479
| : | : | : | : | : | : | : | : | : | : | : | :
Db 442 ITNLG---NTVNNL---SEARSRIEDSDYATEVSNMSRAQ 475

RESULT 4
SLAP CAUCR
ID SLAP CAUCR STANDARD; PRT; 1025 AA.
AC P35828; Q46015; Q9RF12;
DT 01-JUN-1994 (Rel. 29, Created)
DT 16-OCT-2001 (Rel. 40, last sequence update)
DT 28-FEB-2003 (Rel. 41, last annotation update)
DE S-layer protein (Paracrystalline surface layer protein).
GN R5AA OR CC1007.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-21; 646-660 AND 1021-1025.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=93007489; PubMed=1393820;
RA Gilchrist A., Fisher J.A., Smit J.K.;
RT "Nucleotide sequence analysis of the gene encoding the Caulobacter
RL crescentus paracrystalline surface layer protein.";
Can. J. Microbiol. 38:193-202(1992).
RN [2]
RP REVISIONS TO 376; 636 AND 842-843.

RA Awram P.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J53001;
RA Bingle W.H., Awram P.A., Nomenclini J.F., Smit J.K.;
RT "The secretion signal of C. crescentus S-layer protein is located in
RT the C-terminal 82 amino acids of the molecule.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nieman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RT Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
RL [5]
RP SEQUENCE OF 1-313 FROM N.A., AND SEQUENCE OF 1-20.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=89008089; PubMed=3049545;
RA Fisher J.A., Smit J.K., Agabian N.;
RT "Transcriptional analysis of the major surface array gene of
RT Caulobacter crescentus.";
RL J. Bacteriol. 170:4706-4713(1988).
RN [6]
RP CHARACTERIZATION.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=98292737; PubMed=9620954;
RA Awram P., Smit J.K.;
RT "The Caulobacter crescentus paracrystalline S-layer protein is
RT secreted by an ABC transporter (type I) secretion apparatus.";
RL J. Bacteriol. 180:3062-3069(1998).
CC -I- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
CC OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA. PROBABLY ACTS AS A
CC PHYSICAL BARRIER TO PARASITES AND LYTIC ENZYMES.
CC -I- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY. SECRETED BY AN ABC TRANSPORTER
CC (TYPE I) SECRETION APPARATUS.
CC -I- MISCELLANEOUS: THE CALCIUM BINDING OF THIS PROTEIN MAY MEDIATE
CC SURFACE ATTACHMENT BY IONIC BRIDGING BETWEEN THE S-LAYER AND A
CC SPECIFIC MEMBRANE-ASSOCIATED OLIGOSACCHARIDE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AF062345; AAC38665.2; -.

Query Match	5.4%;	Score 130;	DB 1;	Length 1025;
Best Local Similarity	21.0%;	Pred. No. 2.3;		
Matches 101;	Conservative 75;	Mismatches 166;	Indels 140;	Gaps 23;

RESULT 5	
Y866_CHLMU	
ID Y866_CHLMU	STANDARD;
AC Q9PUG1;	PRT; 436 AA

Query Match	5.4%;	Score 129;	DB 1;	Length 436;
Best Local Similarity	20.1%;	Pred. No. 0.9;		
Matches	85;	Conservative	67;	Mismatches 170;
				Indels 100;
				Gaps 14;

```

QY      234 RPSVQGAVDFGVSTAGGLV-ANAGEGDRMLSVQSRDQLRGAFVLAKMKDEPKALLSEET 292
          : | | | | | : | | | |
Db      199 ISGITNIVGFAVSVGGLLSASKSLG-----GLK----SAAFITNET 235

```

```

CC are the predominant cell-type to express the gene. Found
CC abundantly in Purkinje cells.
CC -1- SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.
CC -1- SIMILARITY: Contains 2 calponin-homology (CH) domains.
CC -1- SIMILARITY: Contains 1 PH domain.
CC -1- SIMILARITY: Contains 17 spectrin repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB008551; BAA32699.1; -.
CC DR EMBL; AB001347; BAA32473.1; -.
CC DR EMBL; AF225960; AAC28596.1; -.
CC DR PIR; JE0271; JE0271.
CC DR HSSP; 001082; IBKR.
CC DR GO; GO:0016363; C:nuclear matrix; TAS.
CC DR GO; GO:0008091; C:spectrin; TAS.
CC DR GO; GO:0003779; F:actin binding activity; ISS.
CC DR GO; GO:0030506; F:ankyrin binding activity; ISS.
CC DR GO; GO:0005200; F:structural constituent of cytoskeleton; TAS.
CC DR GO; GO:0017158; P:regulation of calcium ion dependent exocytosis; TAS.
CC DR GO; GO:0016081; P:synaptic vesicle docking; TAS.
CC DR InterPro; IPR001589; Actbind_actinin.
CC DR InterPro; IPR001715; Calponin-like.
CC DR InterPro; IPR001849; PH.
CC DR InterPro; IPR002017; Spectrin.
CC DR InterPro; IPR001605; Spectrin_PH.
CC DR Pfam; PF00307; CH; 2.
CC DR Pfam; PF00169; PH; 1.
CC DR Pfam; PF00435; Spectrin; 17.
CC DR PRINTS; PRO0683; SPECTRINPH.
CC DR SMART; SM00033; CH; 2.
CC DR SMART; SM00233; PH; 1.
CC DR SMART; SM00150; SPEC; 16.
CC DR PROSITE; PS00019; ACTININ_1; 1.
CC DR PROSITE; PS00020; ACTININ_2; 1.
CC DR PROSITE; PS50021; CH; 2.
CC DR PROSITE; PS50003; PH_DOMAIN; 1.
CC KW Cytoskeleton; Membrane; Repeat; Actin-binding; Capping protein.
CC FT DOMAIN 1 278 ACTIN-BINDING (BY SIMILARITY).
CC FT DOMAIN 57 161 CH 1.
CC FT DOMAIN 176 278 CH 2.
CC FT REPEAT 305 415 SPECTRIN 1.
CC FT REPEAT 425 529 SPECTRIN 2.
CC FT REPEAT 531 639 SPECTRIN 3.
CC FT REPEAT 641 745 SPECTRIN 4.
CC FT REPEAT 747 850 SPECTRIN 5.
CC FT REPEAT 852 956 SPECTRIN 6.
CC FT REPEAT 958 1063 SPECTRIN 7.
CC FT REPEAT 1065 1170 SPECTRIN 8.
CC FT REPEAT 1172 1276 SPECTRIN 9.
CC FT REPEAT 1278 1381 SPECTRIN 10.

```


Query Match	5.3%;	Score 128.5;	DB 1;	Length 2388;
Best Local Similarity	21.8%;	Pred. No. 8.4;		
Matches 101;	Conservative 51;	Mismatches 168;	Indels 143;	Gaps 18

```

RESULT 7
CH60_LACAC          STANDARD;      PRT;    546 AA.
ID CH60_LACAC
AC 093G07;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN GROEL OR GROEL.
OS Lactobacillus acidophilus.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1579;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CRL 639;
RA Lorca G.L., Font de Valdez G.;
RT "Characterization of the Lactobacillus acidophilus CRL 639 groESL operon."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: Prevents misfolding and promotes the refolding and proper assembly of unfolded polypeptides generated under stress conditions (By similarity).
CC -I- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of 7 subunits (By similarity).
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -I- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outpost at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF300645; AAK97218.1; -.
DR HAMAP; MF_00600; -.
DR InterPro; IPR001844; Chapernin_Cpn60.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR Pfam; PF00118; cpn60_TCP1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW Chaperone; ATP-binding.
SQ SEQUENCE 546 AA; 58222 MM; CF4F52B2217513F4 CRC64;

Query Match           5.3%; Score 128; DB 1; Length 546;
Best Local Similarity 21.2%; Pred. 1.4;
Matches 111; Conservative 71; Mismatches 196; Indels 146; Gaps 26;

QY 21 ASDASLASSSVRSVTSTCRDIAITDYLKHHVFRAHRFSVI---GSPDERDALA---- 73
   | | : : ||: : : : | : | : | : | : | : | :
Db 2 AKDIKFAENARRSL-----LKGVDKLATDVTKTTIGPKGRNVVLEQSYGNPDITNDGVTI 57
   | | : : | : | : | : | : | : | : | : | : | :
74 -----HNEDIDALVETRA-----NRLVSEGETPATIAETFAKAKEFDRLATVAS-----S 118

```

Db 58 SIELKDHYENMGAKLVAEAAQKTNDIAGDGTATATVLTQAIAREGMQNVTAGANPVGIR 117
QY 119 AFENTPEAASVLYQWOPAIKNGMILA--TPLKPLPLISGALSAMDVG----- 167
Db 118 GIEKATKAAYDELHKISHVESKEQIANVAASASKEVGLIADAMEKVGHGVTTED 177
QY 168 -----TKMDRARGDL-HYLSTSPDKLH-----DAMAVSVKRRHSPALGR 205
Db 178 SRGINTELSVYEGMQFDKGYLSQWMTDNDKMEADLDNPLYILITDKKISNIQDILPLL-Q 236
QY 206 QVNDMGIAVQTFSAINVRTVLAALAS--RPSVQAVDFGVSTAGGLVANAGFGRMLIS 263
Db 237 EIVQOG-----KSLIITADVDVTGEALPTLVINKIRGT--FNVVA-----VKAPGFGDRR-K 284
QY 264 VQSRD--QLRGGAfv---LGMKQKPKALSEETDMLDAYKAIKSASYSGAALNAGKRR 318
Db 285 AQLEDIAALTGGTVITDLDGFEKDKT----- 311
QY 319 GLPLDVATDGLKAVRSLSVATSLSLTKNGLALAGYAGVSK-----LQKMATKNITDSA 370
Db 312 -----IDQLGQARRV---TVTKDSTTIVDG-AGSKEAIKEREDSIRKQIEESTSDFD 359
QY 371 TKAASVQLSNLVGSVGFAGWTTAGLATDPAVKKAESFIQDKVKST-ASSTSYVA---D 426
Db 360 KKKLQERLAELTGGVAVI---HVGATETERELKERRYRIEDALINSTRAAVDEGVVAGG 415
QY 427 QTVKLAKTVKDKSGE-AISSTGASLRSTVNLRHRSAPREADIEE 469
Db 416 ALVDYERKAIKDLKGETSDEQTGINI-----VLRLASAPVRQIAE 454

RESULT 8
NCCB_ALCXX
ID NCCB_ALCXX STANDARD; PRT; 397 AA.
AC Q44585;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Nickel-cobalt-cadmium resistance protein nccb.
GN NCCB.
OS Alcaligenes xylosoxydans xylosoxydans (Achromobacter xylosoxydans).
OG Plasmid pTOM9.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Achromobacter.
OX NCB1_TaxID=515;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=31A;
RX MEDLINE=95050278; PubMed=7961470;
RA Schmidt T., Schlegel H.G.;
RT "Combined nickel-cobalt-cadmium resistance encoded by the ncc locus
of Alcaligenes xylosoxydans 31A."
RL J. Bacteriol. 176:7045-7054(1994).
CC -I- FUNCTION: COMPONENT OF THE NCC CATION-EFFLUX SYSTEM THAT CONFERS
CC RESISTANCE TO NICKEL, COBALT AND CADMIUM.
CC -I- SIMILARITY: HIGH, TO A.EUTROPHUS CNRB AND SOME, TO ALCALIGENES
CC CZCB.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L31363; AAA65105.1; -.
DR PIR; I39579; I39579.
DR InterPro; IPR006143; HlyD.
DR InterPro; IPR005695; Membrane_fus2.
DR Pfam; PF00529; HlyD; 1.
DR TIGRFams; TIGR00999; Ba0102; 1.
KW Plasmid; Nickel; Cobalt; Cadmium; Transport.
SQ SEQUENCE 397 AA; 39933 MW; C6DE57CE314996B8 CRC64;

Query Match 5.3%; Score 127; DB 1; Length 397;
Best Local Similarity 22.8%; Pred. No. 1;
Matches 89; Conservative 59; Mismatches 140; Indels 102; Gaps 19;

QY 149 KPLTPLISGALSAMDQVGTKMDRARG-----DLHYLSTPDKLHDMAVSVKRRHSP 201
Db 9 QPSWPMIAG-VAAALALVG--FGAARGLSPSGAEVSKLAAAEK-----AAASAPAE 60
QY 202 A----LGRQVNDMGIAVQTFSAINVRTVLAALASRPSVQGVDFGVSTAGLVANAGF 257
Db 61 AEVRIPGEYIAAANIAYEPVSAAGVGSVLLAP--ASVAAPGSEBAVIASRAAGV----- 113

QY 258 GDRMLSVQSR--DQLRGAFVLGMKDEPKALSEETDMLDA-----YKAISASYS 307
Db 114 ----LRIQRLGDVAVRAGD-VLALVDSPELAAMAAERVAQAARDLARKTYERESLFFQ 168
QY 308 GAALNAGKMGALPLDVATDGLKAVRSLSVATSLSLTNG--LALAGYAGVSKLQKMATKN 365
Db 169 GVTPRQEMESARIALDVAQAQAEVQRAATVAQAARVSSDGRSVAAVSPiAG-----RTAQ 223

QY 366 ITDSATKAASVQLSNLVGSVGFAGWTTAGLATDPAVKKAESFIQDKVKSTASSTSYVA 425
Db 224 VTIGAYVAPQAELEFRVAGSGAV-----QVEAYV-----TAADTSRIAA 261

QY 426 --DQTVKLAK-----TVKIDSGEALIST-----GASLR--STVN 455
Db 262 GSDATTIVLANGAPLAGRVQAVTPVYSGSARAAVTVVTPVDANSGLIVGEGVQVRILHTKAA 321

QY 456 NLRRRSAPREADIE--EGGISAFSRSETPFQ 483
Db 322 DANAMSVPEDAVQNLGRDVVFVRTOQGFR 351

RESULT 9
CH60_BACHD
ID CH60_BACHD STANDARD; PRT; 544 AA.
AC 050305; Q9KFC3;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)

FT	CONFLICT	500	500	A -> VP (IN REF. 1).
FT	CONFLICT	512	512	MISSING (IN REF. 1).
SC	SEQUENCE	544 AA;	57403 MW;	C769EDB1C903C971 CRC64;
Query Match				
Best Local Similarity 18.9%; Pred. No. 2.7;				
Matches 104; Conservative 75; Mismatches 192; Indels 180; Gaps 23				
QY	21	ASDASLASVSVSVSTTSCRDLOAITDYLK-----HHVFAAHRFSVIGSP-----	65	
Db	2	AKDIKFSDDARSM-----LRGVDKLADAVKVTLPKGRNVYLEKE---GSPPLITNDGVT	54	
QY	66	-----DERDALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATASSAF	120	
Db	55	IAKEIELEDFAENNGAKLVAEVAASKTNDIAGDGTITATV---LAQMIAREGLKNTVSGA-	110	
QY	121	ENTPFAASVLQYQPAINKGDIATPLKPLTPLIS-GALSGAMPDVG---TKMDPARG	176	
Db	111	-NPMVIRKIEKATOVAAVELSKISPIEGKDSIAQVAAISSADDEVGKIIAEAMERVN	169	
QY	177	D-----LHYLSTSPDKLH-----DAMAVSVK	197	
Db	170	DGVITIEESKGFSTELEVEGQKFDGRVASPYMYTDSDKAEAVLDMPPYVLTIDDKTISNIQ	229	
QY	198	RHSPALGRQVDMGIAVQTFSAIENVRTYLAALASRPVSQVAGVDEGVSTAGLVAN---	254	
Db	230	EVLPEVL-EQVVOQGKPI-----LIIAEDVEGALAT-----LVVVKLR	266	
QY	255	-----AGFGDRMLSVQSRQDLRGAFVLGAKKDKPKAALSEETDWDYKAIKSA	304	
Db	267	GTFNAVAVKAPGFGDR-----RKAMLEDI-----	290	
QY	305	SYSGAALNAGKRMA-GLPIDVATDGLKAVRSLVSATSITKNGLALAG-----YAGVS	356	
Db	291	----AIIITGGEVITTEDIGLIDKLSANITQL-GRASKVNVTKENTTIVEGAGESDKIARVN	345	
QY	357	KIOMKATNITDSATKAAVSQLSNLVGSVGFAGTTAGLATDPAYKAESEFIQDKVST	416	
Db	346	QIAQAI EETTSDEVDKEKQERLAKLAGVAVL---KVGATETEKERKURIEDALNST	401	
QY	417	ASSTSYV---ADQTVKLAKTVKDMSGEAISSGCA-----SLRSTVNNLRHRSAPLEADI	467	
Db	402	RAAVEEGIVAGGCTALNVVITKAVSSIGAEDEALGVNITVLALEPEVRQIAPHNAGLEGSV	461	
QY	468	-----EEGG	471	
Db	462	IVERLKKKEAG	472	
RESULT 10				
H1PR_MOUSE				
ID	H1PR_MOUSE	STANDARD;	PRT;	1068 AA.
AC	Q9JKY5;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Huntingtin interacting protein 1 related (Hipl-related).			


```

QY      364KNITDSATKXAVSQTSLNIVGSGVFAGW-----TTAGLATDPVAKKAE-----F 408
Db      677PGFSTSSASLAIN--STVSS--SLAGYSFSTPESSPTSTLTSTVSEAPSTVSSMTTSAPF 732
QY      409IQDKVKSTAS-STTSYVADQTVKLAKTVKDMSGELAISSTGASLSTVANNLRHSAP 463
      1 : : 1 1 1 : : : : 1 1 1 : : 1 1 1 : : 1 1 1
Db      733INNSTARSPPSTASFITESISSIS-SVPLASGDVTSLAA-----HNLITFSAP 781

RESULT 12
MSB2_YEAST
ID      MSB2_YEAST      STANDARD;      PRT;      1306 AA.
AC      P32334;
DT      01-OCT-1993 (Rel. 27, Created)
DT      01-OCT-1993 (Rel. 27, Last sequence update)
DT      01-NOV-1997 (Rel. 35, Last annotation update)
DE      MSB2 protein (Multicopy suppression of a budding defect 2).
GN      MSB2 OR YGR014W.
OS      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX      NCBI_TaxID=4932;
RN      (1)
RP      SEQUENCE FROM N.A.
RX      MEDLINE=92383951; PubMed=1514328;
RA      Bender A., Pringle J.R.;
RT      "A Ser/Thr-rich multicopy suppressor of a cdc24 bud emergence
RT      defect.";
RL      Yeast 8:315-323(1992).
RN      (2)
RP      SEQUENCE FROM N.A.
RC      STRAIN=S288c;
RX      MEDLINE=97435481; PubMed=9290212;
RA      Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
RT      "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
RT      chromosome VII.";
RL      Yeast 13:1077-1090(1997).
CC      -1- PTM: O-GLYCOSYLATED IN THE SER/THR-RICH REGIONS (PROBABLE).
CC      -1- SIMILARITY: SOME, TO YEAST HKR1.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way man-
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M77354; AAA34798.1; -.
DR      EMBL; Z72799; CAA96997.1; -.
DR      PIR; S25370; S25370.
DR      SGD; S0003246; MSB2.
DR      GO; GO:0005887; C:integral to plasma membrane; IMP.
DR      GO; GO:0005034; F:osmosensor activity; IMP.
DR      GO; GO:0000283; P:establishment of cell polarity (sensu Sacch. . .; IGT.
DR      GO; GO:0006970; P:response to osmotic stress; IMP.
RW      Transmembrane; Glycoprotein; Repeat.

```

FT	DOMAIN	698	816	7 X 17 AA	TANDEM REPEATS.
FT	REPEAT	698	714	1.	
FT	REPEAT	715	731	2.	
FT	REPEAT	732	748	3.	
FT	REPEAT	749	765	4.	
FT	REPEAT	766	782	5.	
FT	REPEAT	783	799	6.	
FT	REPEAT	800	816	7.	
SQ	SEQUENCE	1306 AA;	133114 MW;	67D5D984D5CA4A6D	CRC64;
Query Match					
Best Local Similarity 5.0%; Score 120; DB 1; Length 1306;					
Matches 117; Conservativity 20.2%; Pred. No. 12;					
Matches 117; Conservative 79; Mismatches 235; Indels 148; Gaps 24;					
QY	14	AMESFRTASDASLAS-----SSVRSVSTSCRDLOAITDYLKHHVFAAHRFSVIGSPDERD	69		
Db	142	AASSVNAASDVSTASVPIDTSANSIPFTTSNIETTT-----SAPLITSD	185		
QY	70	AALAHNEQIDALVETRANRLYSEGETPATIAETFAKAKEFDRIATTAASSAFENT-PFAAA	128		
Db	186	TPLISTSTMSA-----ADNVFSSANPISASLTTDSSESFDQSTAGALPVQSSADFSS	240		
QY	129	SVLQYMQPAINKGDMLATPLKPLPLISGALSAMDQVG-TKMDRARGDLHYLSTSPDK	187		
Db	241	SEILVQSSAD-----FSSPSSPTTTDIS--LSAAPLOTSSESSFTTASAAPVSTDVVG	293		
QY	188	LHDAMAVSVKRHSPALGRQVDMGIAVQTFPSALNV-----VRT---	225		
Db	294	SSASPVVMMSAAGQIASSSSSTINDPTMSEFSLTSTEVDSDVSTVSALLSAPFLQTSIS	353		
QY	226	----VLAPALASRPSVQGAVDGVSSTAGLVANAGFGD-----RMLS	263		
Db	354	NSFSIVSPSVSFVPS-QSSSDVASSSTANVSSS-FSDIPQSTSGSVSVASASALA	411		
QY	264	VQSRDQLRGATVLMKMKPKALSEETDMLDAYKAIKSASYSGAALNAG-KRMAGLPL	322		
Db	412	FQSTEVYGAASSTMSLSTSLQSTT--LDSSSLASSSASSSDLTDYGVSSASTPL	469		
QY	323	DVATDGLKAVRSL-----VSATSLTKNGLALAGGYAGVSKLQKRA	362		
Db	470	LSASEQASTSSSFVSPSVSFVPSQSSSDVASTAPSVSSSFSTYTSLQAGGSMTPNS	529		
QY	363	TKNI-----TDSATKAAVSOLS-NLVGSVGFAGWTTAGLATPAVKKAESFIQDKVST	416		
Db	530	SSTIVSSSTGSSESAASTASATLSGSSSTY---MAGNLOSQP--PTSSLSIE---SQ	581		
QY	417	ASSTSVVADQTVKLAK---TVKDMSGEA---ISSTGA-----SLRSTVNNLRHRSAPEA	465		
Db	582	ATSTSAVLAASSVSTSTPYTTAGGASTEASSLISSTAETSQVSYSQSTTALQTSFAS	641		
QY	466	DIEEG-----ISAFSRSETPQL	484		
Db	642	STEGSETSSQGFSTSVLVQMPSSISSEFSPQTQM	680		

RESULT 13
HUYA_SERMA

ID	HUYA_SERMA	STANDARD;	PRT; 1608 AA.
AC	P15320;		
DT	01-APR-1990 (Rel. 14, Created)		
DT	01-APR-1990 (Rel. 14, Last sequence update)		
DT	01-NOV-1990 (Rel. 16, Last annotation update)		
DE	Hemolysin precursor.		
GN	SHLA.		
OS	Serratia marcescens.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Serratia.		
OX	NCBI_TaxID=615;		
RN	(1)		
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 31-40.		
RC	STRAIN=SN8;		
RX	MEDLINE=88237037; PubMed=3290200;		
RA	Poole K., Schiebel E., Braun V.;		
RT	"Molecular characterization of the hemolysin determinant of Serratia marcescens.";		
RL	J. Bacteriol. 170:3177-3188(1988).		
CC	-!- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY DEFINED.		
CC	-!- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. SHLA REQUIRES SHIB FUNCTION.		
CC	-!- SUBCELLULAR LOCATION: Outer membrane.		
CC	-!- SIMILARITY: TO P.MIRABILIS HEMOLYSIN (HPMA).		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; M22618; AAA50323.1; -.		
DR	PIR; A28182; A28182.		
KW	Hemolysis; Toxin; Outer membrane; Signal.		
FT	SIGNAL	1	30
FT	CHAIN	31	1608
SQ	SEQUENCE	1608 AA;	165078 MW; D669B476FE7DAD51 CRC64;
Query Match			
Best Local Similarity 5.0%; Score 119.5; DB 1; Length 1608;			
Matches 120; Conservativity 20.5%; Pred. No. 17;			
Matches 120; Conservative 68; Mismatches 242; Indels 155; Gaps 21;			
QY	8	QQPPGVAMESFRTASDASLASSSVRSVSTSCRDLOAITDYLKHHVFAAHRFSVIGSPDE	67
Db	708	QYRAGLRIRHTRDSEKTYRTENSASISLGSVAK-LKAEKDV---TFSGSKLVA-----DK	758
QY	68	RDAALAHNEQIDALVETRANRLYSEGETPATIAET-----FAKAKEFDRIATTAASSAFEN	122
Db	759	GDASVSGN-----KVSFLAADDKTASNTQTKIGGGFYYTGAI DKLGSVGEAGYEN	809
QY	123	TPFAAASVLQYMQPAINKGDMLATPLKPLPLISGALSAMDQVGTMMDRARGDLHY--	180

Db 810 NKTQAQSSKAITS GSDYKGNLTINARDKLTQ-----QGAQHSVGGA YQENAA GVDHLAA 863

QY 181 ---LSTSPDKLHDAMAVSVKRHSPALGRQV---VDMGIAVQTF SALTNVVRTVLAPALASR 234

Db 864 ADTASTTTTKT DVGNIGANVDY SAVTRPERAVGKA AKLDATGVINDI GGI GAPNVGID 923

QY 235 PSVOGAVDFGVSTAGGLVAN---AGFDRMLSVQSRDQ----- 269

Db 924 IGAQGGSSSEKRSSSSQAVVSSVQAGSIDINAKGEVRDQGTQYQASKGAVNLTADSHRSEA 983

QY 270 -----LRGAFVL-----GAKDKEPKAALSEETDMLDAYKAI- 301

Db 984 AANRQDEQSRDTRGSAGRVYTTTGS DLTVDAKGEGGTQRNSSSASQAVTGSID AANGIN 1043

QY 302 ----KSASYSGAALNAGKRM----AG--LP LVDATDGLKAVRSLVGSATSLTKNGLA----- 347

Db 1044 VNVKKDAIYOGTALNGRGKTA VNAAGDIRLDQASDKQESRS GFNVKASAKGFTADSK 1103

QY 348 -LAGGYAGVSKIQKMATQNTDSATKA VSQLS-----NLVGSVG 386

Db 1104 NFGAGEGG-----GTHNGESSSS TAQVGNISQOQVELKAGRDLT LQGTIVKSQGDVS 1156

QY 387 VFAGWTTAGLATDPAVKAESFTQDKVKSTASTTSYVADQTVKLA KTVKMSGEAISST 446

Db 1157 LSAGNKVALQAAESTQTRKESKLSGNIDIGAGSSDS-----KEKTGNLSAG 1203

QY 447 GASLSTVNNLNHRSAPE--ADIEGG---ISAFSRSETFPQLR 485

Db 1204 GAFDIAKVN-----ESATERQGATIASDGKVTLSANGKGDALHLQ 1244

RESULT 14

HS7F_CAEEL STANDARD; PRT; 657 AA.

ID HS7F_CAEEL STANDARD; PRT; 657 AA.

AC P11141; P91135;

DT 01-JUL-1989 (Rel. 11, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE Heat shock 70 kDa protein F, mitochondrial precursor.

GN HSP-6 OR HSP70F OR C37H5.8.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoea; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN 11

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RA Davidson S., Gilliam B.;

RL Submitted (FEB-1997) to the EMBL/Genbank/ DDBJ databases.

RN 12

RP SEQUENCE OF 1-487 FROM N.A.

RC STRAIN=Bristol N2;

RX MEDLINE=89356251; PubMed=2766926;

RA Hoeschl M.F.P., Baillie D.L.;

RT "Characterization of the hsp70 multigene family of Caenorhabditis elegans.";

RL DNA 8:233-243(1989).

CC -1- SUBCELLULAR LOCATION: Mitochondrial.

CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.

CC 1-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; U88315; AAB42371.1; -.

DR EMBL; X07678; CAA30525.1; -.

DR PIR; B32475; B32475.

DR PIR; T25613; T25613.

DR HSP; P04475; 1Dg4.

DR WormPep; C37H5.8; CE08631.

DR InterPro; IPR01023; Hsp70.

DR Pfam; PF00012; HSP70; 1.

DR PRINTS; PRO0301; HEATSHOCK70.

DR ProDom; PD000089; Hsp70; 1.

DR PROSITE; PS00297; HSP70_1; 1.

DR PROSITE; PS00329; HSP70_2; 1.

DR PROSITE; PS01036; HSP70_3; 1.

DR ATP-binding; Heat shock; Multigene family; Mitochondrion;

KW Transit peptide.

KW TRANSIT 1 27 MITOCHONDRION (POTENTIAL).

FT CHAIN 28 657 HEAT SHOCK 70 kDa PROTEIN F.

FT CONFLICT 138 138 V -> E (IN REF. 2).

FT CONFLICT 140 140 S -> P (IN REF. 2).

FT CONFLICT 167 167 A -> P (IN REF. 2).

FT CONFLICT 347 348 KA -> NV (IN REF. 2).

SQ SEQUENCE 657 AA; 70844 MW; 637A1636B6AB836 CRC64;

Query Match 4.9%; Score 119; DB 1; length 657;

Best local Similarity 21.5%; Pred. No. 5.9;

Matches 107; Conservative 60; Mismatches 162; Indels 168; Gaps 26;

QY 29 SSVRSVSTSCRDLOAITDY LKHVFAAHRFSVIGSPDERDALAHNEQIDALVETRANR 88

Db 9 SSARTIARSSLSARSLS DKPKGHVIGID-----LGTNSCVSIMEGTPK 54

QY 89 LYSEGE----TPATIAETFAKAEK-----DRLATYASSAFENTPFAASVL--QYQPA 137

Db 55 VIENAEGVRTTPSTVAFT-ADGERLVGAPAKRQAVTNSA---NTLFATKRLIGRRYEDPE 110

QY 138 INK-----GD-WLATPILKPLTP LISGALSGAMDQVGT KMDRARGDLHYLS 182

Db 111 VQKDLKVVPEYKIVKASNGDAWVEAOGKVVSPQVGAF-----VLMKMKETAES---YLG 161

QY 183 TSPDKLHDAMAVSVKRHSPALGRQV--DMGIAVQTF SALTNVVRTVLAPALASRPSVQGA V 241

Db 162 TTVNN-----AVVTVPAYFENDSQRQATRKDAG-----QISGLNVLRVINEPTFAA----- 204

QY 242 DFGVSTAGGLVANAGFGDRMLSVQSRDQLRGAFVLGKMKDKEPKAALSEETDMLDAYKAI 301

Db 205 ----ALAYGLDKDA--GDKIIAVY----DLGGGTf-----DVSILEIQK-- 238

```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outpost. -
CC CC the European Bioinformatics institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z94723; CAB08123.1; -.
DR EMBL; AL583920; CAC31342.1; -.
DR PIR; C87029; C87029.
DR Leproma; ML0961; -.
DR HAMAP; ME_00172; -: 1.
DR InterPro; IPR006276; Met_syn_B12ind.
DR InterPro; IPR002629; Methionine_synt.
DR Pfam; PF01717; Methionine_synt; 1.
DR ProDom; PD004692; Methionine_synt; 2.
DR TIGRFAMs; TIGR01371; met_syn_B12ind; 1.
KW Transferrase; Methyltransferase; Methionine biosynthesis; Zinc; Repeat;
KW Complete proteome.
FT METAL          647      647       ZINC (BY SIMILARITY) .
FT METAL          649      649       ZINC (BY SIMILARITY) .
FT METAL          732      732       ZINC (BY SIMILARITY) .
SQ SEQUENCE        760 AA;  82234 MW;  1E92FA570B9ACC97 CRC64;

Query Match           4.9%; Score 118.5; DB 1; Length 760;
Best Local Similarity 21.1%; Pred. No. 7.7;
Matches   91; Conservative    66; Mismatches   173; Indels    101; Gaps     18.
```

QY 459 -HRSAPADIE 468
| | | :
Db 394 LHNQGVRAID 404

Search completed: January 6, 2004, 09:45:44
Job time : 21 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2004, 09:36:48 ; Search time 39 Seconds
(without alignments)
3222.349 Million cell updates/sec

Title: US-09-825-414-66
Perfect score: 2407
Sequence: 1 MHINQSAQQPPGVAMESFRT.....EEGGISAFSRSETPFQLRRL 487

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL_23:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_virus:*
 - 16: sp_bacteriap:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8			
Result	Query		
No.	Score	Match Length DB ID	Description

1	1869.5	77.7	486	2	Q9JP34	Q9JP34 pseudomonas
2	1591.5	66.1	484	2	Q8RP03	Q8RP03 pseudomonas
3	460	19.1	518	16	Q8XQF0	Q8XQF0 ralstonia s
4	149.5	6.2	2155	16	Q8UFP9	Q8UFP9 agrobacteri
5	149	6.2	1009	16	Q8U8W1	Q8U8W1 agrobacteri
6	146	6.1	2089	16	Q92K98	Q92K98 rhizobium m
7	143.5	6.0	2055	2	Q85472	Q85472 abiotrophia
8	143	5.9	22152	4	Q8WX17	Q8WX17 homo sapien
9	142	5.9	4753	16	Q8P1I3	Q8P1I3 xanthomonas
10	141	5.9	536	17	Q9HPA2	Q9HPA2 halobacteri
11	140.5	5.8	642	16	Q8XSW6	Q8XSW6 ralstonia s
12	140.5	5.8	1983	16	Q931N9	Q931N9 staphylococ
13	140.5	5.8	2186	16	Q99TB0	Q99TB0 staphylococ
14	140	5.8	503	2	Q9FCQ5	Q9FCQ5 pseudomonas
15	140	5.8	2285	9	O64046	O64046 bacterioph
16	140	5.8	2285	16	O31976	O31976 bacillus su
17	139.5	5.8	491	2	Q93GT1	Q93GT1 campylobact
18	139.5	5.8	1795	16	Q91CJ9	Q91CJ9 staphylococ
19	138.5	5.8	491	2	Q8RTY4	Q8RTY4 campylobact
20	138.5	5.8	548	2	Q9RA74	Q9RA74 streptococc
21	138.5	5.8	3961	16	Q8P942	Q8P942 xanthomonas
22	138	5.7	1242	5	Q9VW05	Q9VW05 drosophila
23	138	5.7	1430	16	Q8P1I2	Q8P1I2 xanthomonas
24	138	5.7	1731	16	Q8U8W4	Q8U8W4 agrobacteri
25	138	5.7	2993	16	Q8NMS0	Q8NMS0 corynebacte
26	137.5	5.7	2283	2	Q8VQ99	Q8VQ99 staphylococ
27	137	5.7	1545	16	Q9RDQ1	Q9RDQ1 streptomyc
28	136	5.7	968	5	Q9VK10	Q9VK10 drosophila
29	135.5	5.6	1952	5	Q8MWV6	Q8MWV6 naegleria g
30	135.5	5.6	2186	16	Q8NM19	Q8NM19 staphylococ
31	135	5.6	973	16	Q8XQ04	Q8XQ04 escherichia
32	135	5.6	6077	12	Q8VAP1	Q8VAP1 white spot
33	135	5.6	6077	12	Q8QTB7	Q8QTB7 white spot
34	134.5	5.6	1307	16	Q8K792	Q8K792 streptococc
35	134.5	5.6	1713	3	Q8TGE1	Q8TGE1 saccharomyc
36	134.5	5.6	2016	5	Q9B1T0	Q9B1T0 plecteurys
37	134.5	5.6	2178	2	Q9KWR3	Q9KWR3 streptococ
38	134	5.6	2535	16	Q8ZDR6	Q8ZDR6 yersinia pe
39	134	5.6	2579	16	Q8DOR8	Q8DOR8 yersinia pe
40	133.5	5.5	1233	16	Q92SD7	Q92SD7 rhizobium m
41	133.5	5.5	1307	16	Q8P1P0	Q8P1P0 streptococ
42	133.5	5.5	2124	16	Q98M03	Q98M03 rhizobium l
43	132.5	5.5	654	16	Q9A4U7	Q9A4U7 caulobacter
44	132	5.5	6077	12	Q91LB5	Q91LB5 white spot
45	131.5	5.5	553	2	Q912M2	Q912M2 rhodobacter

ALIGNMENTS

RESULT 1					
Q9JP34					
ID	Q9JP34	PRELIMINARY;	PRT;	486	AA.
AC	Q9JP34;				
DT	01-OCT-2000	(TEMBLrel. 15, Created)			
DT	01-OCT-2000	(TEMBLrel. 15, last sequence update)			
DT	01-DEC-2001	(TEMBLrel. 19, last annotation update)			

DE	Hypothetical 50.7 kDa protein.
OS	Pseudomonas syringae (pv. tomato).
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC	Pseudomonadaceae; Pseudomonas.
OX	NCBI_TaxID=323;
RN	(1)
RP	SEQUENCE FROM N.A.
RC	STRAIN=DC3000;
RX	MEDLINE=98422476; PubMed=9748456;
RA	Charkowski A.O., Alfano J.R., Preston G., Yuan J., He S.Y.,
RA	Collmer A.;
RT	"The Pseudomonas syringae pv. tomato HrpW protein has domains similar
RT	to harpins and pectate lyases and can elicit the plant hypersensitive
RT	response and bind to pectate.";
RL	J. Bacteriol. 180:5211-5217(1998).
RN	(2)
RP	SEQUENCE FROM N.A.
RC	STRAIN=DC3000;
RX	MEDLINE=20243785; PubMed=10781092;
RA	Alfano J.R., Charkowski A.O., Deng W.L., Badel J.L.,
RA	Petnicki-Ocwieja T., van Dijk K., Collmer A.;
RT	"The Pseudomonas syringae Hrp pathogenicity island has a tripartite
RT	mosaic structure composed of a cluster of type III secretion genes
RT	bounded by exchangeable effector and conserved effector loci that
RT	contribute to parasitic fitness and pathogenicity in plants.";
RL	Proc. Natl. Acad. Sci. U.S.A. 97:4856-4861(2000).
RN	(3)
RP	SEQUENCE FROM N.A.
RC	STRAIN=DC3000;
RA	Ramos A.R., Rehm A.H., Collmer A.R.;
RT	"Pseudomonas syringae pv. tomato DC3000 hrpL through hrcU.";
RL	Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AF232004; AAF71504.1; -.
KW	Hypothetical protein.
SQ	SEQUENCE 486 AA; 50709 MW; 91CFBC03397F8AD1 CRC64;

Query Match					
Best Local Similarity 77.7%; Score 1869.5; DB 2; Length 486;					
Matches 381; Conservative 33; Mismatches 71; Indels 1; Gaps 1;					
QY	1	MHINQSAQQPPGVMESEFRTASDASLASSSVRSVSTTSCRDLQAITDYLGKHFVFAHREFS	60		
Db	1	MHINRRVQQPPVTATDSFRTASDASLASSSVRSVSSDQOREINAIADYLTIDHVFAAHKLP	60		
QY	61	VIGSPDERDALAHNEQIDALIVETFRANRLYSEGETPATIAETFAKAEKEDRLATTASSAF	120		
Db	61	PADSADGQAADVHNAQITALIETFRASRLHFEGETPATIADTFAKAEKLDRLATTSGAL	120		
QY	121	ENTPFAAASVLYVQOPAINKGMLATPPLISGALSGAMDQVGTKMMDRARGDLHY	180		
Db	121	RATPFAMASISLYVQOPAINKGMLPAPLPPLISGALSGAMDQVGTKMMDRATGDLHY	180		
QY	181	LSTSPDKLHDAMAVSVKRHSPLAGROYVDMGIANQTFSAIINVRTVLPALASRPVQGA	240		
Db	181	LSASPDRLHDAMAAVSVKRHSPLARQVLDTGAVQVTSARNAVRTVLPALASRPVQGA	240		
QY	241	VDEGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFLGAKKDKERKALSEETDWLDAYKA	300		

Db 241 VDLGYSMAAGLAAAGCGNRLLSVQSRDHQRGALVTLGLKDKEPKAQLSSENDWLEAYKA 300
QY 301 IKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLSVATSITKNGLALAGYAGVSKIQK 360
Db 301 IKSASYSGAALNAGKRMAGLPLDMATDAGAVRSLSVASSLTQNGLALAGGFAGVGKIQE 360
QY 361 MATKNITDSATKAAVSQSLNLVGSVGFAGWTTAGLATDPAVKKAESFIQDKVKSTAST 420
Db 361 MATKNITDPATKAAVSQSLTNLAGSAAVFAGWTTAALTDPAVKKAESFIQDTVKSTAST 420
QY 421 TSYVADQTVKLAKTVKDMSGEAISSGTASLRSVTNNLRHRSAPeADI EEGISAFSRSST 480
Db 421 TGYVADQTVKLAKTVKDMGGEALTHTGASLRNTVNNLRQRPAREADI EEGTAA-SPSEI 479
QY 481 PFQLRR 486
Db 480 PFRPMR 485

RESULT 2
Q8RP03 PRELIMINARY; PRT; 484 AA.
ID Q8RP03
AC Q8RP03; 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Type III effector HopTox1Pma.
OS Pseudomonas syringae pv. maculicola.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=59511;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ES4326;
RX MEDLINE=21862332; PubMed=11872842;
RA Gutman D.S., Vinatzer B.A., Sarkar S.F., Ranall M.V., Kettler G.,
RA Greenberg J.T.;
RT "A Functional Screen for the Type III (Hrp) Secretome of the Plant
Pathogen Pseudomonas syringae.";
RL Science 295:1722-1726(2002).
DR EMBL; AF458051; AAL84253.1; -.
SQ SEQUENCE 484 AA; 50520 MW; 4DB012DD5688163E CRC64;

Query Match 66.1%; Score 1591.5; DB 2; Length 484;
Best Local Similarity 67.6%; Pred.No. 7.2e-83;
Matches 330; Conservative 51; Mismatches 100; Indels 7; Gaps 3;
QY 1 MHINSAQQPPGVAMESPRTASDASIASSVRSVSTTSCRDLQAITDYLGHHVFAAHRFS 60
Db 1 MYINSISQSSIGTESFHSAQ--SVASSSVRVLSAEQTKLNDITHTYLTDHVFAAHRLP 58
QY 61 VIGSPDERDALAHNEQIDALVETRANRLYSEGTPATIAETFAKAEKFDRLATTASSAF 120
Db 59 VNDINVDDQDVVAHANVEIGKTIISARALRLDDEGTALSIGETFAKAEKFDRLAASGAL 118
QY 121 ENTFFAAASVLYQMOPAINKGDMLATPLKPLTPLISGALSAMDQVGTGMDBRARGDLHY 180

Db 119 RATPFAAASVLYQMOPAINKGDWLPPLKPLAPLVSGALSGVNDQVGTGMNRAATGDLHY 178
QY 181 LSTSPDKLDAMAVSVKRHSPALGRQVDMGIAVQTFSAALNVRTVLADALASRPSVQGA 240
Db 179 LSTAPEKLDHMAASVKRHRPGVNRQAAADLGIAVQTYTARNALRTVLARALASRAVQGA 238
QY 241 VDEGVSTAGLVANAAGFDDMLSVQSRDQLRGAFVLGKDKKEPKAALSEETDMLDAYKA 300
Db 239 VDISVSAAGSLVANAGFSENNHTVQARDHLKGAFLVLCIKDKPKADLSEETDMLDAYRA 298
QY 301 IKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLSVATSITKNGLALAGYAGVSKIQK 360
Db 299 IKSASYSGAALNAGKRVAAGPLDVATDGLKAVRSLSVATSLMONGLVMAAGGFAGVGKIQE 358
QY 361 MATKNITDSATKAAVSQSLNLVGSVGFAGWTTAGLATDPAVKKAESFIQDKVKSTAST 420
Db 359 MATKNITHPATKAAVSQSLTNLAGSAPVFSAWTTAAVAITDPAVKKAESFIQDTVKTTVSN 418
QY 421 TSYVADQTVKLAKTVKDMSGEAISSGTASLRSVTNNLRHRSAPeADI EEGISAFSRS-- 478
Db 419 TSALADKTVKLAKKAGIDASAERIATGASLRDT---LRRTVNEPDI EEGGVAAGTGSPPV 475
QY 479 ETPFQLRR 486
Db 476 AVPFEMR 483

RESULT 3
Q8XQF0 PRELIMINARY; PRT; 518 AA.
ID Q8XQF0
AC Q8XQF0; 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Putative transmembrane protein.
GN RSP1277 OR RS05322.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OG Plasmid megaplasmid.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Ralstoniaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GMI1000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brotier P., Camus J.C., Catolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Sigulier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646083; CAD18428.1; -.
KW Plasmid; Complete proteome.
SQ SEQUENCE 518 AA; 53581 MW; 82D486715F0F60EE CRC64;

```

RA Chapman P., Clendenning J., Deacher G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.-V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323(2001).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houniel K., Gordon J., Vaudin M., Iatchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gurson J., Lomo C., Seear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328(2001).
DR EMBL; AE009096; AAL42354.1; -.
DR EMBL; AE008061; AAK87140.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 2155 AA; 230143 MW; FD28CC6242ADD10 CRC64;

Query Match 6.2%; Score 149.5; DB 16; Length 2155;
Best Local Similarity 20.7%; Pred. No. 9.5;
Matches 96; Conservative 65; Mismatches 176; Indels 127; Gaps 16

QY 69 DAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAKEFDRLATTASSAF----ENTP 124
DB 609 DAFNTHSHAKIDTVLAERSNALF-----GALSASQDFDEALASRSJLAITGSVSGTA 659
QY 125 FAASVYLQYMPARINK-----GDWLATPLKPLTPLISGALSGAMDQVGTKNMDRARGDLH 179
DB 660 EHHAAMLDERRAALNSVADVVERLITETLETRAALITGAVSGIEDRIISDTLESRTAA--- 716
QY 180 YLSTSPDKLHDNAVSVKRRHSPALGRQVYDMGLAVQTFSLALNVVTVLAPLASR----- 234
DB 717 -----LHDVVSGAESRIADTLD-----GRTPALSSAISGVEERIIADMTDSRTLSLD 762
QY 235 -----PSVQGAVDVFGVSTAGGIIVAN-----AGFGDRMLSVQSR--DQLRG 272
DB 763 MTFANVEERLSETLDNRTSALTGTIVASAEKTAGALDSRATFGDVAGAETRIAEITLDG 822
QY 273 -----GAFVLGMRDKEPKKALSEETDMLDAVKAIKASASVGA----- 309
DB 823 RTAALINAVVSGAEER-----IADALDSRTMALDMTFEGAEKEKIAEALDRTALGEL 874
QY 310 ALNAGKRAAGLPIDVATDGLKAVNSLSVATSLTKNGLALAGGYAGVSKLQMKATKNITDS 369
DB 875 VASAEIRIAG-ALDSRTDSLKTV-----VSGAEERIT-DVLDS 910
QY 370 ATKAAVSQLSNLVGSVGVFAGMTAGLATDPAVKKAESFIQDKVNSTASTTSYVADQTV 429

```


Db 911 RTMALDMSFGVEEKITDILDGRTA-----ALKSAVAGVEDRIAGALDSTRALSG--- 961
QY 430 KLAQTVKDMGGEAISSTGASLRSTVNLRHRSAPADIEEGGIS 473
Db 962 -IVSGAEERIAEALDSRTLALDRTISGVEERIAEAMDARASLS 1004

RESULT 5

Q8U8W1 PRELIMINARY; PRT; 1009 AA.
AC Q8U8W1:
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, last annotation update)
DE Ice nucleation protein homolog.
GN INA OR ATU3977 OR AGR_L_1758.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitejima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphitmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Quorillo B., Goldman B.S., Cao Y., Askenazi M., Hailing C., Mullin L.,
RA Houmiel K., Gordon J., Vaudin M., Iatchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328(2001).
DR EMBL; AE009328; AAL44779.1; ALT_INIT.
DR EMBL; AE008285; AAK89449.1; -.
KW Complete proteome.
SQ SEQUENCE 1009 AA; 100468 MW; 14D842830E8C65EE CRC64;

Query Match 6.2%; Score 149; DB 16; Length 1009;
Best Local Similarity 22.4%; Pred. No. 3.4;
Matches 127; Conservative 77; Mismatches 228; Indels 136; Gaps 25;

QY 7 AQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLOAI-1DYLKH---HVFAAHRSVI 62
Db 118 ASQIDANSTDQIALNNSQVAGLSSAQVATLSSDELALFTTDEIKSISANAIAGLSAAAI 177
QY 63 GSPDERDALAHNEQIDALVETRANRLYSEGTPATIAETFAKAKEF-----DELATTA 116
Db 178 AGLSTDNAAALTRKSQIAAMSTQFNALTSGLATFSADVEKASINKI1AGLDVTKLSTGN 237
QY 117 SSAFENTPFAASVILQYNQPAINKGDMLATPPLPLISGALSAMDQVGTIKMDRARG 176
Db 238 IAA1SKAQVSALSTTQFAAMSTDQ-----IKALTSDDQVAGLSSA--QVATLSSD---- 284
QY 177 DLHYLSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIA-----VQ 215
Db 285 ELALFSTDEIKAIKANAVA-GLSAAALALTTDMAALTKTQIAGLSTQUNALTSANLA 343
QY 216 TFSALNVVTVLAPALASRPSVQGVDFGVSTAGGLVA-----NAGFGDRMLSVQSRDQL 270
Db 344 TFSA-DEIKAI1TKALA-----GLDVTKLSTGNIAALTKTQAASLSSSTQFAAMSTDQI 395
QY 271 RG--GAFVLQMKDKEPKALISEE-----TDMLDAYKAIKSASVSGAALNA----- 313
Db 396 KALTSEQVAGLSSAQVATLSSDELALFSTDEISAFSANNVAGLSTAALATGTGNATALT 455
QY 314 GKRMAGL---PLDVATDGLKAVRSIVSATSLTKNGIALAGVAGVSKIQMATKNITDSA 370
Db 456 KTIQAGLSTQUNALTSGLSATFSADVEKAISTK--ALAG-----LDVTKLSTGNVA--AL 507
QY 371 TKAASQSLNVLGVSQVFAQWTT-----AGLATDPAVKKAES----FIQDKVKS 415
Db 508 SKAQVSALSTT-----QFAAMSTDQIKALTSEQVAGLSSAQVATLSSDELALFSTDEIKA 562
QY 416 -----TASSTTSYVADQTVKLAKT-VKDMSGEALISSTGASLRSITVNNLRHRSAP- 464
Db 563 IGANNAVAGLSAALALALTTDNASALTKTQIAGMSSQIINAL-----TSANLATFSADEI 616
QY 465 -----ADIEEGGISAFSRSE 479
Db 617 KAITTKALGGLDVTKLSTGNIAALTKAQ 644

RESULT 6

Q92K98 PRELIMINARY; PRT; 2089 AA.
AC Q92K98;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)
DE Hypothetical transmembrane protein SMC00190.
GN R01816 OR SMC00190.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;

Query Match	6.1%;	Score 146;	DB 16;	Length 2089;
Best Local Similarity	18.9%;	Pred. No. 14;		
Matches 106;	Conservative 84;	Mismatches 208;	Indels 162;	Gaps 17;

```

Dd      4362  ITSNTFTDSTNSTPSTSLIMIESSGVTITQTQPTGALTQGPYLLDFTSTWPLYETPLAV  4421

QY      100  AETFAKAKEKEDRLA-----TTASSAFEN-----TFFAASVLYQYMPAIN--KGDW  143
          |::||:  ::  |:  |  |  |||  ::  ||:  :|
Dd      4422  TPDEMQSEKTTLLISKQPKDVTWTSPPSVAETSIFFSLPLFLVYTI----PPATSTLQGH  4477
          |::||:  ::  |:  |  |  |||  ::  ||:  :|

QY      144  LATPLKPLPLISGALSGAMDQVGTAKMDRAGDLHYLSTSPDKLHD-----AMAV  194
          ::||:  ::||:  ::  |  |  :  ::||  ||:  :|
Dd      4478  TSSPVSATSVLTSGLV-----KTTDMINTSMEPVTNSPQINNPSEI LATLAATT  4528
          ::||:  ::||:  ::  |  |  :  ::||  ||:  :|

QY      195  SVKRHSPALGRQVDMGIAVOTFSALNVRTYLAPALASRPSVOAGVDFGVSTAGIIVAN  254
          ::  |:::  |::||  |  |:::  |  ::||  :|
Dd      4529  DIETIHPSINKAVTNNG-----TASSAHLHSTL--PVSESPS-----TATSPWPA  4573
          ::  |:::  |::||  |  |:::  |  ::||  :|

QY      255  AGEGDRMLSVQSRDQLRGAFVLGMKDKEPKALSEETDMLDAYKAIKSASYSGAALNAG  314
          :  ||:  |::  :  |  ||  ||  ||
Dd      4574  SSMGDALASIS-----IPG-----SETTD-----  4592
          :  ||:  |::  :  |  ||  ||  ||

```

Dd 4593 --IEGPTSLTAGKKNSTLDQEMNSTTESNIIILSNVSGAITEATKMEVPSFDAFIPT 4650

Oy 362 ---ATK-----NITDSATKAASQL-----SNLVGSGV-- 386

D5 4651 PAQSTKEPDI FSVASSRLSNBPMTISTHMTTQTGSSGATSKIPLALDITSTLETSAGTP 4710

D6 387 --VEAGW-----TAGLATDPAVKKALESFIQDKKSTAS-----S 419

Qy 420 TTSVADQTVKLAKTVK-----DMSGELISSTIGASLRSTVNILHRSAPEADIE 468
 | : | : : | : | : : | : : : : | : ||
Db 4771 TSSPVSMSVLTSSLVKTAGKVDTSETVTSSPQSMSTLDDISVTSAAFTDIE 4824

RESULT 9	
Q8PLT3	
ID Q8PLI3	PRELIMINARY;
AC Q8PLT3,	
DT 01-OCT-2002 (TREMBlrel, 22, Created)	PRT; 4753 AA.

DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)
DE Filamentous haemagglutinin.
GN FHAB OR XAC1815.
OS Xanthomonas axonopodis (pv. citri).
ON
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;

```

OX      NCBI_TaxID=92829;
RN      [1]
RP      SEQUENCE FROM N.A.

```

RX MEDLINE=22022145; PubMed=12024217;
da Silva A.C.R., Ferro J.A., Reinech F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteloro-Vitcorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergro F., Clapiņa L.P.,
Cicarrelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorfy H.,

RESULT	9
Q8PLI3	
ID	Q8PLI3
AC	Q8PLI3;
DT	01-OCT-2002 (TREMblrel. 22, Created)
DT	01-OCT-2002 (TREMblrel. 22, last sequence update)
DT	01-MAR-2003 (TREMblrel. 23, last annotation update)
DE	Filamentous haemagglutinin.
GN	FHAB OR XAC1815.
OS	Xanthomonas axonopodis (pv. citri).
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC	Xanthomonadaceae; Xanthomonas.
OX	NCBI_TaxID=92829;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=306 / ATCC 13902 / XV 101;
RX	MEDLINE=22022145; PubMed=12024217;
RA	da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furian L.R.,
RA	Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA	Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA	Cardozo G., Camaran F., Cardozo J., Chambergo F., Ciapina L.P.,
RA	Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,


```

Db          1448 AGTDA5LHGN-----RIATGRDVLKSSG 1471

RESULT 10
Q9HPA2
ID          Q9HPA2          PRELIMINARY;          PRT;          536 AA.
AC          Q9HPA2;
DT          01-MAR-2001 (TREMBLrel. 16, Created)
DT          01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT          01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE          Htr17 transducer.
DE          HTRI17 OR VNGI733G.
OS          Halobacterium sp. (strain NRC-1).
OS          Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC          Halobacteriaceae; Halobacterium.
OX          NCBI_TaxID=64091;
RN          (1)
RP          SEQUENCE FROM N.A.

```

	Query Match	5.9%; Score 141; DB 17; Length 536;
	Best Local Similarity	23.2%; Pred. No. 3.9;
	Matches 119; Conservative	78; Mismatches 197; Indels 120; Gaps 23;
QY	12 GVAMESFRTASDASIASSVSRSVSTSCRDLOAITDYLGHHVFAAHRSFVIGSPDERDAA	71
	: :: : : : : : : :	
Dd	63 GPVMDAFST----SLLAVLIGVAISSGTISTIVED-LKTSNARHR-----RRKDAEQAK	113
QY	72 LAHNEDIDALIVETRANRLYSSEGETPATIAETFAKAKEFDRIATTAASAAGENTPFAAASVL	131
	: : : : : : : : : : : :	
Dd	114 RRADDEQRQAQGEAKAEAQDRAAREEAETLAKLGERAERAQSLSKASEYSVMKKAS--	171
QY	132 QYNQPAINKGDMLATPLKLPTPLI-SCALSICAMDQVG-----TKAADRAR-----	175
	: : : : : : : : :	
Dd	172 -----GD-----LTVRANSSSDSDAWTEVGQAFENEMTAAFEENVARTREFAGD	214

[illegible]

RESULT 12	
Q931N9	
ID Q931N9	PRELIMINARY; PRT; 1983 AA.
AC Q931N9;	
DT 01-DEC-2001	(TrEMBLrel. 19, Created)
DT 01-DEC-2001	(TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003	(TrEMBLrel. 23, Last annotation update)
DE MRP protein.	

Query Match	5.88;	Score 140.5;	DB 16;	Length 1983;
Best Local Similarity	19.38;	Pred. No. 28;		
Matches 105;	Conservative 94;	Mismatches 203;	Indels 143;	Gaps 21;

1 250 NAEVDQATVAENNI DAVQVDVKKQARDKIT-----AEVAKRIEAVKQTPNATDEEKQ 1304

1405 SÖVNÖ 1405

SQ SEQUENCE 2186 AA; 238448 MW; 0CC84836F1868A4F CRC64;

QY 76 EQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAFENTPEAASVLQYMQ 135

Db 973 --IDKSAETQNTHTIFNNND--ATLEEQAAQQLDDQAVATAKQININ-----AADTNQGEVA 1023
 QY 136 PAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMPDRARGDLHYLSTSPDKLHDAMAVS 195
 Db 1024 QAKDQGTQNTIVVIQPAIQVKTDBANA-----VNDKAREAITNTINATPGATREEKEGA 1075
 QY 196 VKRHPALGRQVVDMGIAVQTFSAIINVRT-----VLAPALASRPSVQGA----- 240
 Db 1076 INRVNTLKNRALTDIGV-TSTIAMVNSIRDDAVNQIGAVQPHVTKKQTATGVLNDLATAK 1134
 QY 241 -----VDEGVSTAGGLVANAGFGDRMLSVQSRDQLRGGA----- 275
 Db 1135 KOEINQNTNATTEEKQVALNQVQDELATA---INNINQADTNAEVDAQQLGTKALINAIQ 1191
 QY 276 -----VLGKKDKPEKPAALSEETDMLDAYKAIKSASYSGAALNAGKMA----- 318
 Db 1192 PNIVKKPAAIAQINQHVNAKLAIEINATPDATNDEKNAAIN--TINQDRQQAIESIKQANT 1249
 QY 319 GLPLD---VATDGLKAVRSLSVATSLTKNGLALAGYAGVSKLQKMA--TKNITDSATK 372
 Db 1250 NAEVDQAATVAENINIDAVQVDVVKQAARDKIT-----AEVAKRIEAVKQTPNATDEEKQ 1304
 QY 373 AAVSQLSNLVGSVGFAGWTTAGLATDPAYKKA-ESFIQDKVKSTASSTSY---ADQ 427
 Db 1305 AAVNQINQL-----KDQAINQINQNTNDQVDTTNOAVNALDNEVEAV 1348
 QY 428 TVK-----LAKTVKDSGEAISSTGASLSTVANNLRHRSAP EADIEGGISAFSRSETP 481
 Db 1349 VIKPKAIAIDIEKAVKEKQOOIID---SLDSTDNEKEVASQALAKEKEKALAAIDQAOITN 1404
 QY 482 FQLRR 486
 Db 1405 SQVNO 1409
 RESULT 14
 Q9FCQ5 PRELIMINARY; PRT; 503 AA.
 AC Q9FCQ5;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Flagellin (Fragment).
 GN FLIC.
 OS Pseudomonas fragi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=296;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCIMB8542;
 RA Bellingham N.F., Morgan J.A.W., Saunders J.R., Winstanley C.;
 RT "Flagellin gene sequence phylogeny in the genus Pseudomonas.";
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ297534; CAC03723.1; -.
 DR InterPro; IPR001492; FlagellinN.

DR InterPro; IPR001029; Flagellin_C.
 DR Pfam; PF00700; Flagellin_C; 1.
 DR Pfam; PF00669; Flagellin_N; 1.
 DR PRINTS; PR00207; FLAGELLIN.
 DR ProDom; PD000316; Flagellin_C; 2.
 FT NON_TER 1 1
 FT NON_TER 503 503
 SQ SEQUENCE 503 AA; 51085 MW; AE42EEFB521BF7C1 CRC64;

Query Match 5.8%; Score 140; DB 2; Length 503;
 Best Local Similarity 24.3%; Pred. No. 4.1;
 Matches 99; Conservative 43; Mismatches 135; Indels 130; Gaps 19;

QY 147 PLKPLTPLISGA-----LSGAMDQVGTKMPDRARGDLHYLSTSPDKLHDAMAVSVK 197
 Db 10 PVKGMAVTIANANNSSIAQTAEGAM-QESTNIIQRLR-ELAIQSANGDKSADDRASLQQ 67
 QY 198 RHSPALG-----ROVDMGIAVOTF----- 217
 Db 68 EFTAKVGEITRISSTTTFGSRNLLDGSFQVQADANQITISFGMSDISATGLKSGYG 127
 QY 218 --SALNVRTVLAPALASRPSVQGAVDFGVSTAGLVANAGFGDRMLSVQSRDQLRGGA 275
 Db 128 EASAAAGVSTLSANV-----VGANDATFKATGAALFAAVDQTLTINGTD----- 174
 QY 276 VLGKKDK-----PKAALSEETDMLDAYKAIKSASYSGA 309
 Db 175 IATAKDSKIADVAVEINKQTSKTVTASADAATGTKLTLSASSTFAVGSAAADGAFVAA 234
 QY 310 ALNAGKRAAGPLDVATDGLK--AVRSLV-SATSLRK-NGLALAGYAGVSKLQKMATK 364
 Db 235 ATPA-KNLIGRPVEIQVNGVKVTIAAGSLENAAAAITKANTDSKTVNASVKKDGRVLVTS 293
 QY 365 -----NITDSATKAAVSQLSNLVGSVGFAGWTTAGLATDPAY-----KKAESFI 409
 Db 294 EKQQAINILADSTGTGPGSLSKL-----GLTAGSTQAKLTNDTSVSFNGVGKFKGDS-- 347
 QY 410 QDKVKSTAS-----TTSVYADQTVKLAKTVKDSGEAISSTGASLR 451
 Db 348 MDIVSSINSASTGVTASKNADNTLKLFT-KD-----ITTAGGSCR 388

RESULT 15
 O64046 PRELIMINARY; PRT; 2285 AA.
 AC O64046;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Putative transglycosylase.
 GN YOMI.
 OS Bacteriophage SPBc2.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
 OX NCBI_TaxID=66797;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lazarevic V., Duesterhoeft A., Soldo B., Hilbert H., Manuel C.,

RA Karamata D.;
RT "The complete nucleotide sequence of the Bacillus subtilis SPbetac2
RT prophage.";
RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF020713; AAC13005.1; -.
DR HSSP; P00718; 153L.
DR InterPro; IPR002886; Peptidase_M37.
DR InterPro; IPR000189; SLT_domain.
DR Pfam; PF01551; Peptidase_M37; 1.
DR Pfam; PF01464; SLT; 1.
DR PROSITE; PS00922; TRANSGLYCOSYLASE; 1.
SQ SEQUENCE 2285 AA; 252310 MW; FF602C227754B357 CRC64;

Query Match 5.8%; Score 140; DB 9; Length 2285;
Best Local Similarity 20.5%; Pred.No. 36;
Matches 128; Conservative 80; Mismatches 224; Indels 192; Gaps 26;

QY 15 MESFRTASDASIASSSVRSVSTTSCRDLOAITDYLKHHVFAHRSV-IGSPDERDALA 73
:|::: : : : | | : : ||: ||| | : : | | | : | :
Db 251 IELYQROAVNQNLNTRYGSSMSSNRQAVQDYLN---AVNSLVSTGSNNIRSQIQS 306
QY 74 HNEQIDAL---VETRANRLYSEGETPATIAET-----AKA 106
| | | : | ||: | | | : : ||
Db 307 LMQFRELASNAQTAANQASSFG---AELTQTFKSMSTYLLISGLFYGAISGLKEMVSOA 363
QY 107 EKFDRLATTSASFENTPFAAASVLQYMQPAINKGDLATPLKLTPLISGALSAMDV 166
: | | | : : : | : | | | : : : : : | :
Db 364 IEIDTLMTNIRRVNPEPDYKYNEL---QESI DLGDTLSNKITDILOMTGDFGRMGFDES 420
QY 167 GTRKMDRARGDLHYLS-TSPDKLDAMAVSVKRHPALGRV-----VDMGIAVQTF 217
: : : | : | : || : : : : : : : | | |
Db 421 ELSTLTQAVLQNVSDLPDDTVNLTITAAVLNFINIANDSISADKINEVDNNYAVTTL 480
QY 218 SALNVVRTLAPALASRPSVQAVDFGVSTA-----GGLVAN-----AGFGRMLS 263
| : | | : | | : | : | | : | | : | : | : | :
Db 481 DLANSIRK--AGSTASTFGVELNDLIGYTTALASTRESGNI VGNLSLKTIFARI GNNQSS 538
QY 264 VQSRDQLRGAFVLGMKDKPEKALISE---ETDWL-DAYKAIKSASYSG----- 308
::: ||: | : | : || : | | | | | : | :
Db 539 IKALEQIGISVKTAGGEAKSASDLISEVAGKMDTISDAQKQNTSIGVAGIYQLSRFNAM 598
QY 309 ----AALNAGK----- 315
| | | |
Db 599 NNFSTIAQNAKTAANSTGSAMSEQKYADSLQARVNKLQNNFTFEFALASDAFISDGLIE 658
QY 316 --RMAGLP L DVAI DGLKAV---RSLVSATS-----LTKNGLALAAG-YAGVSKL--QKVA 362
: | | : : | : | | : : | | : | | | : : : : | :
Db 659 FTQAAAGSLINASTGVIKSVGFLLPPLAAVSTATLLSKNTRTLASSLILGTRANGQETLA 718
QY 363 TKNITDSATKAAVSQ-----LSNLVGSVGVFAGWTTAGLATDPA-VKKA----- 405
| : | : ||| : | ||| | : | | | |
Db 719 TAGLEAGWTRAAVASRVLKTALRGILVSTLVGGAFAALGWALESLISSFAEAKKAKDPE 778
QY 406 -----ESFIQ-----DKVSTASTTSYVADQ---TVKLAKT----VK 436
: | | : | | | | : | : : : | : | : | |
Db 779 QSQQTNVEAITTNKSDTDKL IQYKELQKVESRSLTSDEQEVLYQVYQQLAQTFPALVK 838

QY 437 --DMSGEAISSTGASLRSTVNNLR 458
| | | | | : | :
Db 839 GYDSQGNAILKTNKELEKAIENTK 862

Search completed: January 6, 2004, 09:45:15
Job time : 43 secs